Chapter One

Cones of Hermitian matrices and trigonometric polynomials

In this chapter we study cones in the real Hilbert spaces of Hermitian matrices and real valued trigonometric polynomials. Based on an approach using such cones and their duals, we establish various extension results for positive semidefinite matrices and nonnegative trigonometric polynomials. In addition, we show the connection with semidefinite programming and include some numerical experiments.

1.1 CONES AND THEIR BASIC PROPERTIES

Let $\mathcal{H}$ be a Hilbert space over $\mathbb{R}$ with inner product $\langle \cdot, \cdot \rangle$. A nonempty subset $\mathcal{C}$ of $\mathcal{H}$ is called a *cone* if

(i) $\mathcal{C} + \mathcal{C} \subseteq \mathcal{C}$,

(ii) $\alpha\mathcal{C} \subseteq \mathcal{C}$ for all $\alpha > 0$.

Note that a cone is *convex* (i.e., if $C_1, C_2 \in \mathcal{C}$ then $sC_1 + (1-s)C_2 \in \mathcal{C}$ for $s \in [0,1]$). The cone $\mathcal{C}$ is *closed* if $\mathcal{C} = \overline{\mathcal{C}}$, where $\overline{\mathcal{C}}$ denotes the closure of $\mathcal{C}$ (in the topology induced by the norm $\|\cdot\|$ on $\mathcal{H}$, where, as usual, $\|\cdot\| = \langle \cdot, \cdot \rangle^{1/2}$). For example, the set $\{(x,y) : 0 \leq x \leq y\}$ is a closed cone in $\mathbb{R}^2$. This cone is the intersection of the cones $\{(x,y) : |x| \leq y\}$ and $\{(x,y) : x, y \geq 0\}$. In general, intersections and sums of cones are again cones:

Proposition 1.1.1 *Let $\mathcal{H}$ be a Hilbert space and let $\mathcal{C}_1$ and $\mathcal{C}_2$ be cones in $\mathcal{H}$. Then the following are also cones:*

(i) $\mathcal{C}_1 \cap \mathcal{C}_2$, *and*

(ii) $\mathcal{C}_1 + \mathcal{C}_2 = \{C_1 + C_2 \ : \ C_1 \in \mathcal{C}_1, C_2 \in \mathcal{C}_2\}$.

Proof. The proof is left to the reader (see Exercise 1.6.1). □

Given a cone $\mathcal{C}$ the *dual* $\mathcal{C}^*$ of $\mathcal{C}$ is defined via

$$\mathcal{C}^* = \{L \in \mathcal{H} \ : \ \langle L, K \rangle \geq 0 \text{ for all } K \in \mathcal{C}\}.$$

The notion of a dual cone is of special importance in optimization problems, as optimality conditions are naturally formulated using the dual cone. As a

simple example, it is not hard to see that the dual of $\{(x,y) : 0 \le x \le y\}$ is the cone $\{(x,y) : x \ge 0 \text{ and } y \ge -x\}$. We call a cone $\mathcal{C}$ *selfdual* if $\mathcal{C}^* = \mathcal{C}$. For example, $\{(x,y) : x, y \ge 0\}$ is selfdual.

Lemma 1.1.2 *The dual of a cone is again a cone, which happens to be closed. Any subspace W of $\mathcal{H}$ is a cone, and its dual is equal to its orthogonal complement (i.e., if W is a subspace then $W^* = W^\perp := \{h \in \mathcal{H} : \langle h, w\rangle = 0 \text{ for all } w \in W\}$). Moreover, for cones $\mathcal{C}$, $\mathcal{C}_1$, and $\mathcal{C}_2$ we have that*

(i) *if $\mathcal{C}_1 \subseteq \mathcal{C}_2$ then $\mathcal{C}_2^* \subseteq \mathcal{C}_1^*$;*

(ii) $(\mathcal{C}^*)^* = \overline{\mathcal{C}}$;

(iii) $\mathcal{C}_1^* + \mathcal{C}_2^* \subseteq (\mathcal{C}_1 \cap \mathcal{C}_2)^*$;

(iv) $(\mathcal{C}_1 + \mathcal{C}_2)^* = \mathcal{C}_1^* \cap \mathcal{C}_2^*$.

Proof. The proof of this proposition is left as an exercise (see Exercise 1.6.3). □

An *extreme ray of a cone* $\mathcal{C}$ is a subset of $\mathcal{C} \cup \{0\}$ of the form $\{\alpha K \ : \alpha \ge 0\}$, where $0 \ne K \in \mathcal{C}$ is such that

$$K = A + B,\ A, B \in \mathcal{C} \Rightarrow A = \alpha K \text{ for some } \alpha \in [0, \infty).$$

The cone $\{(x,y) : 0 \le x \le y\}$ has the extreme rays $\{(x,0) : x \ge 0\}$ and $\{(x,x) : x \ge 0\}$.

There are two cones that we are particularly interested in: the cone PSD of positive semidefinite matrices and the cone $\mathbb{T}\mathrm{Pol}^+$ of trigonometric polynomials that take on nonnegative values on the unit circle. We will study some of the basics of these cones in this section, and in the remainder of this chapter we will study related cones. The cone $\mathbb{R}\mathrm{Pol}^+$ of polynomials that are nonnegative on the real line is of interest as well. Some of its properties will be developed in the exercises.

1.1.1 The cone PSD_n

Let $\mathbb{C}^{n\times n}$ be the Hilbert space over $\mathbb{C}$ of $n \times n$ matrices endowed with the inner product

$$\langle A, B\rangle = \mathrm{tr}(AB^*),$$

where $\mathrm{tr}(M)$ denotes the trace of the square matrix M and B^* denotes the complex conjugate transpose of the matrix B. Consider the subset of Hermitian matrices $\mathcal{H}_n = \{A \in \mathbb{C}^{n\times n} \ : \ A = A^*\}$, which itself is a Hilbert space over $\mathbb{R}$ under the same inner product $\langle A, B\rangle = \mathrm{tr}(AB^*) = \mathrm{tr}(AB)$. When we restrict ourselves to real matrices we will write $\mathcal{H}_{n,\mathbb{R}} = \mathcal{H}_n \cap \mathbb{R}^{n\times n}$. The following results on the cone PSD_n of positive semidefinite matrices are well known.

Lemma 1.1.3 *The set PSD_n of all $n \times n$ positive semidefinite matrices is a selfdual cone in the Hilbert space $\mathcal{H}_n$.*

Proof. Let A and B be $n \times n$ positive semidefinite matrices. We denote by $A^{\frac{1}{2}}$ the unique positive semidefinite matrix the square of which is A. Then, by a well-known property of the trace, $\operatorname{tr}(AB) = \operatorname{tr}(A^{\frac{1}{2}}BA^{\frac{1}{2}}) \geq 0$. Thus $\mathrm{PSD}_n \subseteq (\mathrm{PSD}_n)^*$.

For the converse, let $A \in \mathcal{H}_n$ be such that $\operatorname{tr}(AB) \geq 0$ for every positive semidefinite matrix B. For all $v \in \mathbb{C}^n$, the rank 1 matrix $B = vv^*$ is positive semidefinite, thus $0 \leq \operatorname{tr}(Avv^*) = \langle Av, v\rangle$, so A is positive semidefinite. □

Proposition 1.1.4 *The extreme rays of* PSD_n *are given by* $\{\alpha vv^* : \alpha \geq 0\}$*, where* v *is a nonzero vector in* $\mathbb{C}^n$*. In other words, all extreme rays of* PSD_n *are generated by rank 1 matrices.*

Proof. Let $v \in \mathbb{C}^n$ and suppose that $vv^* = A + B$ with A and B positive semidefinite. If $w \in \mathbb{C}^n$ is orthogonal to v then we get that $0 = w^*vv^*w = w^*Aw + w^*Bw$, and as both A and B are positive semidefinite we get that $w^*Aw = 0 = w^*Bw$. Thus $\|A^{\frac{1}{2}}w\| = 0 = \|B^{\frac{1}{2}}w\|$. This yields that $A^{\frac{1}{2}}w = 0 = B^{\frac{1}{2}}w$, and thus $Aw = 0 = Bw$. As this holds for all w orthogonal to v, we get that A and B have at most rank 1. If they have rank 1, the eigenvector corresponding to the single nonzero eigenvalue must be a multiple of v, and this implies that both A and B are a multiple of vv^*.

Conversely, if $K \in \mathrm{PSD}_n$ write $K = v_1v_1^* + \cdots + v_kv_k^*$, where $v_1, \ldots, v_k$ are nonzero vectors in $\mathbb{C}^n$ and k is the rank of K (use, for instance, a *Cholesky factorization* $K = LL^*$, and let $v_1, \ldots, v_k$ be the nonzero columns of the Cholesky factor L). But then it follows immediately that if $k \geq 2$, the matrix K does not generate an extreme ray of PSD_n. □

1.1.2 The cone $\mathbb{T}\mathrm{Pol}_n^+$

We consider *trigonometric polynomials* $p(z) = \sum_{k=-n}^{n} p_k z^k$ with the inner product

$$\langle p, g\rangle = \frac{1}{2\pi}\int_0^{2\pi} p(e^{i\theta})\overline{g(e^{i\theta})}d\theta = \sum_{k=-n}^{n} p_k\overline{g_k},$$

where $g(z) = \sum_{k=-n}^{n} g_k z^k$. We are particularly interested in the Hilbert space $\mathbb{T}\mathrm{Pol}_n$ over $\mathbb{R}$ consisting of those trigonometric polynomials $p(z)$ that are real valued on the unit circle, that is, $p \in \mathbb{T}\mathrm{Pol}_n$ if and only if p is of degree $\leq n$ and $p(z) \in \mathbb{R}$ for $z \in \mathbb{T}$. It is not hard to see that the latter condition is equivalent to $p_k = \overline{p_{-k}}$, $k = 0, \ldots, n$. For such real-valued trigonometric polynomials we have that $\langle p, g\rangle = \sum_{k=-n}^{n} p_k g_{-k}$. The cone that we are interested in is formed by those trigonometric polynomials that are nonnegative on the unit circle:

$$\mathbb{T}\mathrm{Pol}_n^+ = \{p \in \mathbb{T}\mathrm{Pol}_n : p(z) \geq 0 \text{ for all } z \in \mathbb{T}\}.$$

A crucial property of these nonnegative trigonometric polynomials is that they factorize as the modulus squared of a single polynomial. This property stated in the next theorem is known as the *Fejér-Riesz factorization* property; it is in fact the key to determining the dual of $\mathbb{T}\mathrm{Pol}_n^+$ and its extreme

rays. We call a polynomial q *outer* if all its roots are outside the open unit disk; that is, q is outer when $q(z) \neq 0$ for $z \in \mathbb{D}$. The degree n polynomial $q(z) = \sum_{k=0}^{n} q_n z^n$ is called *co-outer* if all its roots are inside the closed unit disk and $q_n \neq 0$ (in some contexts it is convenient to interpret $q_n \neq 0$ to mean that ∞ is not a root of $q(z)$). One easily sees that $q(z)$ is outer if and only if $z^n \overline{q(1/\overline{z})}$ is co-outer.

Theorem 1.1.5 *The trigonometric polynomial $p(z) = \sum_{k=-n}^{n} p_k z^k$, $p_n \neq 0$, is nonnegative on the unit circle $\mathbb{T}$ if and only if there exists a polynomial $q(z) = \sum_{k=0}^{n} q_k z^k$ such that*

$$p(z) = q(z)\overline{q(1/\overline{z})},\ z \in \mathbb{C},\ z \neq 0.$$

In particular, $p(z) = |q(z)|^2$ for all $z \in \mathbb{T}$. The polynomial q may be chosen to be outer (co-outer) and in that case q is unique up to a scalar factor of modulus 1.

When the factor q is outer (co-outer), we refer to the equality $p = |q|^2$ as an *outer (co-outer) factorization.* To make the outer factorization unique one can insist that $q_0 \geq 0$, in which case one can refer to q as *the* outer factor of p. Similarly, to make the co-outer factorization unique one can insist that $q_n \geq 0$, in which case one can refer to q as *the* co-outer factor of p.

Proof. The "if" statement is trivial, so we focus on the "only if" statement. So let $p \in \mathbb{T}\mathrm{Pol}_n^+$. Using that $p_{-k} = \overline{p}_k$ for $k = 0, \ldots, n$, we see that the polynomial

$$g(z) = z^n p(z) = \overline{p}_n + \cdots + p_0 z^n + \cdots + p_n z^{2n}$$

satisfies the relation

$$g(z) = z^{2n}\overline{g(1/\overline{z})},\ z \neq 0. \tag{1.1.1}$$

Let $z_1, \ldots, z_{2n}$ be the set of all zeros of g counted according to their multiplicities. Since all z_j are different from zero, it follows from (1.1.1) that $1/\overline{z}_j$ is also a zero of g of the same multiplicity as z_j. We prove that if $z_j \in \mathbb{T}$, in which case $z_j = 1/\overline{z}_j$, then $z_j = e^{it_j}$ has even multiplicity. The function $\phi : \mathbb{R} \to \mathbb{C}$ defined by $\phi(t) = p(e^{it})$ is differentiable and nonnegative on $\mathbb{R}$. This implies that the multiplicity of t_j as a zero of ϕ is even since ϕ has a local minimum at t_j, and consequently z_j is an even multiplicity zero of g. So we can renumber the zeros of g so that $z_1, \ldots, z_n, 1/\overline{z}_1, \ldots, 1/\overline{z}_n$ represent the $2n$ zeros of g. Note that we can choose $z_1, \ldots, z_n$ so that $|z_k| \geq 1$, $k = 1, \ldots, n$ (or, if we prefer, so that $|z_k| \leq 1$, $k = 1, \ldots, n$). Then we have

$$g(z) = p_n \prod_{j=1}^{n} (z - z_j) \prod_{j=1}^{n} \left(z - \frac{1}{\overline{z}_j} \right),$$

so

$$p(z) = z^{-n} g(z) = p_n \prod_{j=1}^{n} (z - z_j) \prod_{j=1}^{n} \left(1 - \frac{1}{z\overline{z}_j} \right) \tag{1.1.2}$$

$$= c\prod_{j=1}^{n}(z-z_j)\prod_{j=1}^{n}\left(\frac{1}{z}-\overline{z}_j\right),$$

where $c = p_n(\prod_{j=1}^{n}(-\overline{z}_j))^{-1}$. Since p is nonnegative on $\mathbb{T}$ and positive somewhere, at $\alpha \in \mathbb{T}$, say, we get that $c > 0$ (as $c = \frac{\prod_j |\alpha - z_j|^2}{p(\alpha)}$). With $q(z) = \sqrt{c}\prod_{j=1}^{n}(z-z_j)$, (1.1.2) implies that $p(z) = q(z)\overline{q(1/\overline{z})}$. Note that choosing $|z_k| \geq 1$, $k = 1, \ldots, n$, leads to an outer factorization, and that choosing $|z_k| \leq 1$, $k = 1, \ldots, n$, leads to a co-outer factorization. The uniqueness of the (co-)outer factor up to a scalar factor of modulus 1 is left as an exercise (see Exercise 1.6.5). □

To obtain a description of the dual cone of $\mathbb{T}\mathrm{Pol}_n^+$, it will be of use to rewrite the inner product $\langle p, g\rangle$ in terms of the factorization $p = |q|^2$. For this purpose, introduce the *Toeplitz matrix*

$$T_g := (g_{i-j})_{i,j=0}^{n} = \begin{pmatrix} g_0 & g_{-1} & \cdots & g_{-n} \\ g_1 & g_0 & \cdots & g_{-n+1} \\ \vdots & \ddots & \ddots & \vdots \\ g_n & g_{n-1} & \cdots & g_0 \end{pmatrix},$$

where $g(z) = \sum_{k=-n}^{n} g_k z^k$. It is now a straightforward calculation to see that $p(z) = |q_0 + \cdots + q_n z^n|^2$, $z \in \mathbb{T}$, yields

$$\langle p, g\rangle = \begin{pmatrix}\overline{q_0} & \cdots & \overline{q_n}\end{pmatrix}\begin{pmatrix} g_0 & g_1 & \cdots & g_{-n} \\ g_1 & g_0 & \cdots & g_{-n+1} \\ \vdots & \ddots & \ddots & \vdots \\ g_n & g_{n-1} & \cdots & g_0 \end{pmatrix}\begin{pmatrix} q_0 \\ \vdots \\ q_n \end{pmatrix} = x^* T_g x, \tag{1.1.3}$$

where $x = \begin{pmatrix} q_0 & \cdots & q_n \end{pmatrix}^T$ and the superscript T denotes taking the transpose. From this it is straightforward to see that the dual cone of $\mathbb{T}\mathrm{Pol}_n^+$ consists exactly of those trigonometric polynomials g for which the associated Toeplitz matrix T_g is positive semidefinite (notation: $T_g \geq 0$).

Lemma 1.1.6 *The dual cone of* $\mathbb{T}\mathrm{Pol}_n^+$ *is given by*

$$(\mathbb{T}\mathrm{Pol}_n^+)^* = \{g \in \mathbb{T}\mathrm{Pol}_n : T_g \geq 0\}. \tag{1.1.4}$$

Proof. Clearly, if $p = |q|^2 \in \mathbb{T}\mathrm{Pol}_n^+$ and $g \in \mathbb{T}\mathrm{Pol}_n$ is such that $T_g \geq 0$, we obtain by (1.1.3) that $\langle p, g\rangle \geq 0$. This gives $\supseteq$ in (1.1.4). For the converse, suppose that $g \in \mathbb{T}\mathrm{Pol}_n$ is such that T_g is not positive semidefinite. Then choose $x = \begin{pmatrix} q_0 & \cdots & q_n \end{pmatrix}^T$ such that $x^* T_g x < 0$, and let $p \in \mathbb{T}\mathrm{Pol}_n$ be given via $p(z) = |q_0 + \cdots + q_n z^n|^2$, $z \in \mathbb{T}$. Then (1.1.3) yields that $\langle p, g\rangle < 0$, and thus $g \notin (\mathbb{T}\mathrm{Pol}_n^+)^*$. This yields $\subseteq$ in (1.1.4). □

Next, we show that the extreme rays of $\mathbb{T}\mathrm{Pol}_n^+$ are generated exactly by those nonnegative trigonometric polynomials that have $2n$ roots (counting multiplicity) on the unit circle.

Proposition 1.1.7 *The trigonometric polynomial* $q(z) = \sum_{k=-n}^{n} q_k z^k$ *in* $\mathbb{T}\mathrm{Pol}_n^+$ *generates an extreme ray of* $\mathbb{T}\mathrm{Pol}_n^+$ *if and only if* $q_n \neq 0$ *and all the roots of* q *are on the unit circle; equivalently,* $q(z)$ *is of the form*

$$q(z) = c \prod_{k=1}^{n} (z - e^{it_k}) \left(\frac{1}{z} - e^{-it_k} \right), \tag{1.1.5}$$

for some $c > 0$ *and* $t_k \in \mathbb{R}$, $k = 1, \ldots, n$.

Proof. First suppose that $q(z)$ is of the form (1.1.5) and let $q = g + h$ with $g, h \in \mathbb{T}\mathrm{Pol}_n^+$. As $q(e^{it_k}) = 0$, $k = 1, \ldots, n$, and g and h are nonnegative on the unit circle, we get that $g(e^{it_k}) = h(e^{it_k}) = 0$, $k = 1, \ldots, n$. Notice that $q \geq g$ on $\mathbb{T}$ implies that the multiplicity of a root on $\mathbb{T}$ of g must be at least as large as the multiplicity of the same root of q (use the construction of the function ϕ in the proof of Theorem 1.1.5). Using this one now sees that g and h must be multiples of q.

Next, suppose that q has fewer that $2n$ roots on the unit circle (counting multiplicity). Using the construction in the proof of Theorem 1.1.5 we can factor q as $q = gh$, where $g \in \mathbb{T}\mathrm{Pol}_m^+$, $m < n$, has all its roots on the unit circle, $h \in \mathbb{T}\mathrm{Pol}_k^+$ has none of its roots on the unit circle, and $k + m \leq n$. If h is not a constant, we can subtract a positive constant ϵ, say, from h while remaining in $\mathbb{T}\mathrm{Pol}_k^+$ (e.g., take $\epsilon = \min_{z \in \mathbb{T}} h(z) > 0$). Then $q = \epsilon g + (h - \epsilon) g$ gives a way of writing q as a sum of elements in $\mathbb{T}\mathrm{Pol}_n^+$ that are not scalar multiples of q (as h was not a constant function). In case $h \equiv c$ is a constant function, we can write

$$q(z) = g(z)h(z) = g(z) \left(\frac{1}{2}c - \delta z - \frac{\delta}{z} \right) + g(z) \left(\frac{1}{2}c + \delta z + \frac{\delta}{z} \right),$$

which for $0 < \delta < \frac{c}{4}$ gives a way of writing q as a sum of elements in $\mathbb{T}\mathrm{Pol}_{m+1}^+ \subseteq \mathbb{T}\mathrm{Pol}_n^+$ that are not scalar multiples of q. This proves that q does not generate an extreme ray of $\mathbb{T}\mathrm{Pol}_n^+$. □

Notice that if we consider the cone of nonnegative trigonometric polynomials without any restriction on the degree, the argument in the proof of Proposition 1.1.7 shows that none of its elements generate an extreme ray. In other words, the cone $\mathbb{T}\mathrm{Pol}^+ = \bigcup_{n=0}^{\infty} \mathbb{T}\mathrm{Pol}_n^+$ does not have any extreme rays.

1.2 CONES OF HERMITIAN MATRICES

A natural cone to study is the cone of positive semidefinite matrices that have zeros in certain fixed locations, that is, positive semidefinite matrices with a given sparsity pattern. As we shall see, the dual of this cone consists of those Hermitian matrices which can be made into a positive semidefinite matrix by changing the entries in the given locations. As such, this dual cone is related to the positive semidefinite completion problem. While in this section we analyze this problem from the viewpoint of cones and their

properties, we shall directly investigate the completion problem in detail in the next chapter, then in the context of operator matrices.

For presenting the results of this section we first need some graph theoretical preliminaries. An *undirected graph* is a pair $G = (V, E)$ in which V, the *vertex set*, is a finite set and the *edge set* E is a symmetric binary relation on V such that $(v, v) \notin E$ for all $v \in V$. The *adjacency set* of a vertex v is denoted by $\mathrm{Adj}(v)$, that is, $w \in \mathrm{Adj}(v)$ if $(v, w) \in E$. Given a subset $A \subseteq V$, define the *subgraph induced* by A by $G|A = (A, E|A)$, in which $E|A = \{(x, y) \in E \ : \ x \in A \text{ and } y \in A\}$. The *complete graph* is a graph with the property that every pair of distinct vertices is adjacent. A subset $K \subseteq V$ is a *clique* if the induced graph $G|K$ on K is complete. A clique is called a *maximal clique* if it is not a proper subset of another clique.

A subset $P \subseteq \{1, \ldots, n\} \times \{1, \ldots, n\}$ with the properties

(i) $(i, i) \in P$ for $i = 1, \ldots, n$,

(ii) $(i, j) \in P$ if and only if $(j, i) \in P$

is called an $n \times n$ *symmetric pattern*. Such a pattern is said to be a *sparsity pattern* (location of required zeros) for a matrix $A \in \mathcal{H}_n$, if every $1 \leq i, j \leq n$ such that $(i, j) \notin P$ implies that the (i, j) and (j, i) entries of A are 0.

With an $n \times n$ symmetric pattern P we associate the undirected graph $G = (V, E)$, with $V = \{1, \ldots, n\}$ and $(i, j) \in E$ if and only if $(i, j) \in P$ and $i \neq j$. For a pattern with associated graph $G = (V, E)$, we introduce the following subspace of $\mathcal{H}_n$:

$$\mathcal{H}_G = \{A \in \mathcal{H}_n : A_{ij} = 0 \text{ for all } (i, j) \notin P\},$$

that is, the set of all $n \times n$ Hermitian matrices with sparsity pattern P. It is easy to see that

$$\mathcal{H}_G^{\perp} = \{A \in \mathcal{H}_n : A_{ij} = 0 \text{ for all } (i, j) \in P\}.$$

When A is an $n \times n$ matrix $A = (A_{ij})_{i,j=1}^n$ and $K \subseteq \{1, \ldots, n\}$, then $A|K$ denotes the $\mathrm{card}K \times \mathrm{card}K$ principal submatrix

$$A|K = (A_{ij})_{i,j \in K} = (A_{ij})_{(i,j) \in K \times K}.$$

Here $\mathrm{card}K$ denotes the cardinality of the set K.

We now define the following cones in $\mathcal{H}_G$:

$$\begin{aligned}
\mathrm{PPSD}_G &= \{\, A \in \mathcal{H}_G : A|K \geq 0 \text{ for all cliques } K \text{ of } G \,\}, \\
\mathrm{PSD}_G &= \{\, X \in \mathcal{H}_G : X \geq 0 \,\}, \\
\mathcal{A}_G &= \{\, Y \in \mathcal{H}_G : \text{ there is a } W \in \mathcal{H}_G^{\perp} \text{ such that } Y + W \geq 0 \,\}, \\
\mathcal{B}_G &= \{\, B \in \mathcal{H}_G : B = \textstyle\sum_{i=1}^{n_B} B_i \text{ where} \\
&\qquad\qquad B_1, \ldots, B_{n_B} \in \mathrm{PSD}_G \text{ are of rank } 1 \,\}.
\end{aligned}$$

The abbreviation PPSD stands for partially positive semidefinite.

We have the following descriptions of the duals of PPSD_G and PSD_G. For Hermitian matrices we introduce the Loewner order $\leq$, defined by $A \leq B$ if and only if $B - A \geq 0$ (i.e., $B - A$ is positive semidefinite).

Proposition 1.2.1 *Let $G = (V, E)$ be a graph. The cones PPSD_G, PSD_G, $\mathcal{A}_G$, and $\mathcal{B}_G$ are closed, and their duals (in $\mathcal{H}_G$) are identified as*

$$(\mathrm{PSD}_G)^* = \mathcal{A}_G, \text{ and } (\mathrm{PPSD}_G)^* = \mathcal{B}_G.$$

Moreover,

$$(\mathrm{PPSD}_G)^* \subseteq \mathrm{PSD}_G \subseteq (\mathrm{PSD}_G)^* \subseteq \mathrm{PPSD}_G. \tag{1.2.1}$$

Proof. The closedness of PPSD_G, $\mathcal{A}_G$, and PSD_G is trivial. For the closedness of $\mathcal{B}_G$, first observe that if $J_1, \ldots, J_p$ are all the maximal cliques of G, then any $B \in \mathcal{B}_G$ can be written as $B = B_1 + \cdots + B_p$, where $B_k \geq 0$ and the nonzero entries of B_k lie in $J_k \times J_k$, $k = 1, \ldots, p$. Note that $B_k \leq B$, $k = 1, \ldots, p$. Let now $A^{(n)} = \sum_{k=1}^{p} A_k^{(n)} \in \mathcal{B}_G$ be so that the nonzero entries of $A_k^{(n)}$ lie in $J_k \times J_k$, $k = 1, \ldots, p$, and assume that $A^{(n)}$ converges to A as $n \to \infty$. We need to show that $A \in \mathcal{B}_G$. Then $A_1^{(n)}$ is a bounded sequence of matrices, and thus there is a subsequence $\{A_1^{(n_k)}\}_{k\in\mathbb{N}}$, that converges to A_1, say. Note that A_1 is positive semidefinite and has only nonzero entries in $J_1 \times J_1$. Next take a subsequence $\{A_2^{(n_{k_l})}\}_{l\in\mathbb{N}}$, of $\{A_2^{(n_k)}\}_{k\in\mathbb{N}}$ that converges to A_2, say, which is automatically positive semidefinite and has only nonzero entries in $J_2 \times J_2$. Repeating this argument, we ultimately obtain $m_1 < m_2 < \cdots$ so that $\lim_{j\to\infty} A_k^{(m_j)} = A_k \geq 0$ with A_k having only nonzero entries in $J_k \times J_k$. But then

$$A = \lim_{n\to\infty} A^{(n)} = \sum_{k=1}^{p} \lim_{j\to\infty} A_k^{(m_j)} = A_1 + \cdots + A_p \in \mathcal{B}_G,$$

proving the closedness of $\mathcal{B}_G$.

By Lemma 1.1.2 it is true that if $\mathcal{C}$ is a cone and W is a subspace, then $(\mathcal{C} \cap W)^* = \mathcal{C}^* + W^\perp$. This implies the duality between PSD_G and $\mathcal{A}_G$ since $\mathrm{PSD}_G = \mathrm{PSD}_n \cap \mathcal{H}_G$ and PSD_n is selfdual.

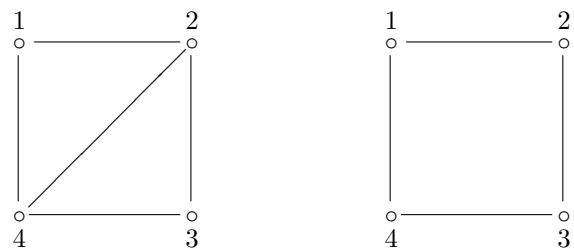
For the duality of PPSD_G and $\mathcal{B}_G$ note that if $X \in \mathrm{PSD}_G$ and rank $X = 1$, then $X = ww^*$, where w is a vector with support in a clique of G. This shows that any element in PPSD_G lies in the dual of $\mathcal{B}_G$. Next if $A \notin \mathrm{PPSD}_G$, then there is a clique K such that $A|K$ is not positive semidefinite. Thus there is a vector v so that $v^*(A|K)v < 0$. But then one can pad the matrix vv^* with zeros, and obtain a positive semidefinite rank 1 matrix B with all its nonzero entries in the clique K, so that $\langle A, B\rangle < 0$. As $B \in \mathcal{B}_G$, this shows that A is not in the dual of $\mathcal{B}_G$. □

It is not hard to see that equality between PSD_G and $(\mathrm{PSD}_G)^*$ holds only in case the maximal cliques of G are disjoint, and in that case all four cones are equal.

In order to explore when the cones PSD_G^* and PPSD_G are equal, it is convenient to introduce the following graph theoretic notions. A *path* $[v_1, \ldots, v_k]$ in a graph $G = (V, E)$ is a sequence of vertices such that $(v_j, v_{j+1}) \in E$ for $j = 1, \ldots, k-1$. The path $[v_1, \ldots, v_k]$ is referred to as a path between v_1 and v_k. A graph is called *connected* if there exists a path between any two different vertices in the graph. A *cycle* of length $k > 2$ is a path $[v_1, \ldots, v_k, v_1]$

in which $v_1, \ldots, v_k$ are distinct. A graph G is called *chordal* if every cycle of length greater than 3 possesses a *chord*, that is, an edge joining two nonconsecutive vertices of the cycle.

Below are two graphs and the corresponding symmetric patterns (pictured as partial matrices). The graph on the left is chordal, while the one on the right (the four cycle) is not.



$$\begin{pmatrix} * & * & ? & * \\ * & * & * & * \\ ? & * & * & * \\ * & * & * & * \end{pmatrix}, \qquad \begin{pmatrix} * & * & ? & * \\ * & * & * & ? \\ ? & * & * & * \\ * & ? & * & * \end{pmatrix}.$$

An ordering $\sigma = [v_1, \ldots, v_n]$ of the vertices of a graph is called a *perfect vertex elimination scheme (or perfect scheme)* if each set

$$S_i = \{v_j \in \mathrm{Adj}(v_i) \ : \ j > i\} \tag{1.2.2}$$

is a clique. A vertex v of G is said to be *simplicial* when $\mathrm{Adj}(v)$ is a clique. Thus $\sigma = [v_1, \ldots, v_n]$ is a perfect scheme if each v_i is simplicial in the induced graph $G|\{v_i, \ldots, v_n\}$.

A subset $S \subset V$ is an $a-b$ *vertex separator* for nonadjacent vertices a and b if the removal of S from the graph separates a and b into distinct connected components. If no proper subset of S is an $a-b$ separator, then S is called a *minimal $a-b$ vertex separator.*

Proposition 1.2.2 *Every minimal vertex separator of a chordal graph is complete.*

Proof. Let S be a minimal $a-b$ separator in a chordal graph $G = (V, E)$, and let A and B be the connected components in $G|V - S$ containing a and b, respectively. Let $x, y \in S$. Each vertex in S must be connected to at least one vertex in A and at least one in B. We can choose minimal length paths $[x, a_1, \ldots, a_r, y]$ and $[y, b_1, \ldots, b_s, x]$, such that $a_i \in A$ for $i = 1, \ldots, r$ and $b_j \in B$ for $j = 1, \ldots, s$. Then $[x, a_1, \ldots, a_r, y, b_1, \ldots, b_s, x]$ is a cycle in G of length at least 4. Chordality of G implies it must have a chord. Since $(a_i, b_j) \notin E$ by the definition of a vertex separator, $(a_i, a_k) \notin E$ and $(b_k, b_l) \notin E$ by the minimality of r and s, the only possible chord is (x, y). Thus S is a clique. $\square$

The next result is known as Dirac's lemma.

Lemma 1.2.3 *Every chordal graph G has a simplicial vertex, and if G is not a clique, then it has two nonadjacent simplicial vertices.*

Proof. We proceed by induction on n, the number of vertices of G. When $n = 1$ or $n = 2$, the result is trivial. Assume $n \geq 3$ and the result is true for graphs with fewer than n vertices. Assume $G = (V, E)$ has n vertices. If G is complete, the result holds. Assume G is not complete, and let S be a minimal $a - b$ separator for two nonadjacent vertices a and b. Let A and B be the connected components of a respectively b in $G|V - S$.

By our assumption, either the subgraph $G|A \cup S$ has two nonadjacent simplicial vertices one of which must be in A (since by Proposition 1.2.2, $G|S$ is complete), or $G|(A \cup S)$ is complete and any vertex in A is simplicial in $G|A \cup S$. Since $\mathrm{Adj}(A) \subseteq A \cup S$, a simplicial vertex of $G|A \cup S$ in A is simplicial in G. Similarly, B contains a simplicial vertex of G, and this proves the lemma. □

The following result is an algorithmic characterization of chordal graphs.

Theorem 1.2.4 *An undirected graph is chordal if and only if it has a perfect scheme. Moreover, any simplicial vertex can start a perfect scheme.*

Proof. Assume $G = (V, E)$ be a chordal graph with n vertices. Assume every chordal graph with fewer than n vertices has a perfect scheme. (For $n = 1$ the result is trivial.) By Lemma 1.2.3, G has a simplicial vertex x. Let $[v_1, \ldots, v_{n-1}]$ be a perfect scheme for $G|V - \{x\}$. Then $[x, v_1, \ldots, v_{n-1}]$ is a perfect scheme for G.

Let $G = (V, E)$ be a graph with a perfect scheme $\sigma = [v_1, \ldots, v_n]$ and assume C is a cycle of length at least 4 in G. Let x be the vertex of C with the smallest index in σ. By the definition of a perfect scheme, the two vertices in C adjacent to x must be connected by an edge, so C has a chord. □

The following result is an important consequence of Theorem 1.2.4. We will use the following lemma.

Lemma 1.2.5 *Let $A > 0$. Then $\left(\begin{smallmatrix} A & B \\ B^* & C \end{smallmatrix}\right)$ is positive (semi)definite if and only if $C - B^*A^{-1}B$ is.*

Proof. This follows immediately from the following factorization:

$$\begin{pmatrix} A & B \\ B^* & C \end{pmatrix} = \begin{pmatrix} I & 0 \\ B^*A^{-1} & I \end{pmatrix} \begin{pmatrix} A & 0 \\ 0 & C - B^*A^{-1}B \end{pmatrix} \begin{pmatrix} I & A^{-1}B \\ 0 & I \end{pmatrix}. \qquad \square$$

In Section 2.4 we will see that $C - B^*A^{-1}B$ is a so-called "Schur complement," and that more general versions of this lemma hold.

Proposition 1.2.6 *Let $A \in \mathrm{PSD}_G$, where G is a chordal graph with n vertices. Then A can be written as $A = \sum_{i=1}^{r} w_i w_i^*$, where $r = \mathrm{rank} A$ and $w_i \in \mathbb{C}^n$ are nonzero vectors such that $w_i w_i^* \in \mathrm{PSD}_G$ for $i = 1, \ldots, r$. In particular, we have $\mathrm{PSD}_G = (\mathrm{PPSD}_G)^*$.*

Proof. We use induction on n. For $n = 1$ the result is trivial and we assume it holds for $n-1$. Let $A \in \mathrm{PSD}_G$, where $G = (V, E)$ is a chordal graph with n vertices, and let $r = \mathrm{rank} A$. We can assume without loss of generality that the vertex 1 is simplicial in G (otherwise we reorder the rows and columns of A in an order that starts with a simplicial vertex). If $a_{11} = 0$, then the first row and column of A are zero, so the result follows from the assumption for $n-1$. Otherwise, let $w_1 \in \mathbb{C}^n$ be the first column of A multiplied by $\frac{1}{\sqrt{a_{11}}}$. An entry (k, l), $2 \le k, l \le n$ of $w_1 w_1^*$ is nonzero if $(1, k)$ and $(1, l)$ are in E. Since the vertex 1 is simplicial, we have $(k.l) \in E$. Then, by Lemma 1.2.5, $A - w_1 w_1^*$ is a positive semidefinite matrix of rank $r-1$, and has its first row and column equal to zero, and $(A - w_1 w_1^*)|\{2, \ldots, n\} \in \mathrm{PSD}_{G|\{2,\ldots,n\}}$. Since $G|\{2, \ldots, n\}$ is also chordal, by our assumption for $n-1$, we have that $A - w_1 w_1^* = \sum_{i=2}^{r} w_i w_i^*$, where each $w_i \in \mathbb{C}^n$ is a nonzero vector with its first component equal to 0. This completes our proof. □

Proposition 1.2.6 has two immediate corollaries.

Corollary 1.2.7 *Let P be a symmetric pattern with associated graph $G = (V, E)$. Then Gaussian elimination can be carried out on every $A \in \mathrm{PSD}_G$ such that in the process no entry corresponding to $(i, j) \notin P$ is changed even temporarily to a nonzero, if and only if $\sigma = [1, 2, \ldots, n]$ is a perfect scheme for G.*

Corollary 1.2.8 *Let $G = (V, E)$ be a graph. Then the lower-upper Cholesky factorization $A = LL^*$ of every $A \in \mathrm{PSD}_G$ satisfies $L_{ij} = 0$ for $1 \le j < i \le n$ such that $(i, j) \notin E$, if and only if $\sigma = [1, 2, \ldots, n]$ is a perfect scheme for G.*

Proof. Follows immediately from the fact that $L = \begin{pmatrix} w_1 & w_2 & \cdots w_n \end{pmatrix}$, where $w_1, \ldots, w_n$ are obtained recursively as in the proof of Proposition 1.2.6. □

Proposition 1.2.9 *Let $G = (V, E)$ be a nonchordal graph. Then $(\mathrm{PSD}_G)^*$ is a proper subset of PPSD_G.*

Proof. For $m \ge 4$ define the $m \times m$ Toeplitz matrix

$$A_m = \begin{pmatrix} 1 & 1 & 0 & 0 & \cdots & -1 \\ 1 & 1 & 1 & 0 & \cdots & 0 \\ 0 & 1 & 1 & 1 & \cdots & 0 \\ \vdots & \ddots & \ddots & \ddots & \ddots & \vdots \\ -1 & 0 & 0 & 0 & \cdots & 1 \end{pmatrix}, \tag{1.2.3}$$

the graph of which is the chordless cycle $[1, \ldots, m, 1]$. Each A_m is partially positive semidefinite since both matrices $\left(\begin{smallmatrix} 1 & 1 \\ 1 & 1 \end{smallmatrix}\right)$ and $\left(\begin{smallmatrix} 1 & -1 \\ -1 & 1 \end{smallmatrix}\right)$ are positive semidefinite. We cannot modify the zeros of A_m in any way to obtain a positive semidefinite matrix, since the only positive semidefinite matrix with the three middle diagonals equal to 1 is the matrix of all 1's.

Let $G = (V, E)$, $V = \{1, \ldots, n\}$, be a nonchordal graph and assume without loss of generality that it contains the chordless cycle $[1, \ldots, m, 1]$, $4 \leq m \leq n$. Let A be the matrix with graph G having 1 on its main diagonal, the matrix A_m in (1.2.3) as its $m \times m$ upper-left corner, and 0 on any other position. Then $A \in \mathrm{PPSD}_G$, but $A \notin (PSD_G)^*$, since the zeros in its upper left $m \times m$ corner cannot be modified such that this corner becomes a positive semidefinite matrix. □

The following result summarizes the above results and shows that equality between $(\mathrm{PPSD}_G)^*$ and PSD_G (or equivalently, between $(\mathrm{PSD}_G)^*$ and PPSD_G) occurs exactly when G is chordal.

Theorem 1.2.10 *Let $G = (V, E)$ be a graph. Then the following are equivalent.*

(i) *G is chordal.*

(ii) $\mathrm{PPSD}_G = (\mathrm{PSD}_G)^*$.

(iii) $(\mathrm{PPSD}_G)^* = \mathrm{PSD}_G$.

(iv) *There exists a permutation σ of $[1, \ldots, n]$ such that after reordering the rows and columns of every $A \in \mathrm{PSD}_G$ by the order σ, A has the lower-upper Cholesky factorization $A = LL^*$ with $L_{ij} = 0$ for every $1 \leq i < j \leq n$ such that $(i, j) \notin E$.*

(v) *The only matrices that generate extreme rays of PSD_G are rank 1 matrices.*

Proof. (i) ⇒ (iv) follows from Corollary 1.2.8, (iv) ⇒ (iii) from Corollary 1.2.8 and Proposition 1.2.6, (iii) ⇒ (ii) from Lemma 1.1.2 and the closedness of the cones, while (ii) ⇒ (i) follows from Proposition 1.2.9. The implications (i) ⇒ (iii) ⇒ (v) follow from Proposition 1.2.6. □

Let us give an example of a higher rank matrix that generates an extreme ray in the nonchordal case.

Example 1.2.11 Let G be given by the graph representing the chordless cycle $[1, 2, 3, 4, 1]$, and let

$$K = \begin{pmatrix} 1 & 1 & 0 & 1 \\ 1 & 2 & 1 & 0 \\ 0 & 1 & 1 & -1 \\ 1 & 0 & -1 & 2 \end{pmatrix}.$$

Then K generates an extreme ray for PSD_G. Indeed, suppose that $K = A + B$ with $A, B \in \mathrm{PSD}_G$. Notice that if we let

$$V = \begin{pmatrix} 1 & 1 & 0 & 1 \\ 0 & 1 & 1 & -1 \end{pmatrix} = \begin{pmatrix} v_1 & v_2 & v_3 & v_4 \end{pmatrix},$$

then $K = V^*V$. As $K = A + B$ and $A, B \geq 0$, it follows that the nullspace of K lies in the nullspaces of both A and B. Using this, it is not hard to see that A and B must be of the form

$$A = V^*\widetilde{A}V,\ B = V^*\widetilde{B}V,$$

where $\widetilde{A}$ and $\widetilde{B}$ are 2×2 positive semidefinite matrices. As $A, B \in \mathrm{PSD}_G$, we have that

$$v_1^*\widetilde{A}v_3 = 0 = v_1^*\widetilde{B}v_3 \ \text{ and } \ v_2^*\widetilde{A}v_4 = 0 = v_2^*\widetilde{B}v_4.$$

But then both $\widetilde{A}$ and $\widetilde{B}$ are orthogonal to the matrices

$$v_3v_1^* = \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix},\ v_1v_3^* = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix},$$

$$v_2v_4^* = \begin{pmatrix} 1 & -1 \\ 1 & -1 \end{pmatrix},\ v_4v_2^* = \begin{pmatrix} 1 & 1 \\ -1 & -1 \end{pmatrix}.$$

But then $\widetilde{A}$ and $\widetilde{B}$ must be multiples of the identity, and thus A and B are multiples of K.

1.3 CONES OF TRIGONOMETRIC POLYNOMIALS

Similarly to Section 1.2, we introduce in this section four cones, which now contain real-valued trigonometric polynomials. We establish the relationship between these cones, identify their duals, and discuss the situation when they are pairwise equal.

Let $d \geq 1$ and let S be a finite subset of $\mathbb{Z}^d$. We consider the vector space Pol_S of trigonometric polynomials p with Fourier support in S, that is, $p(z) = \sum_{\lambda \in S} p_\lambda z^\lambda$, where $z = (e^{it_1}, \ldots, e^{it_d})$, $\lambda = (\lambda_1, \ldots, \lambda_d)$, and $z^\lambda = (e^{i\lambda_1 t_1}, \ldots, e^{i\lambda_d t_d})$. Note that in the notation Pol_S we do not highlight the number of variables; it is implicit in the set S. In subsequent notation we have also chosen not to attach the number of variables, as it would make the notation more cumbersome. We hope that this does not lead to confusion. The complex numbers $\{p_\lambda\}_{\lambda \in S}$ are referred to as the *(Fourier) coefficients* of p. The *Fourier support* of p is given by $\{\lambda \ : \ p_\lambda \neq 0\}$. We endow Pol_S with the inner product

$$\langle p, q \rangle = \frac{1}{(2\pi)^d} \int\limits_{[0,2\pi]^d} p(e^{it_1}, \ldots, e^{it_d})\overline{q(e^{it_1}, \ldots, e^{it_d})}dt_1 \ldots dt_d = \sum_{\lambda \in S} p_\lambda \overline{q_\lambda},$$

where $\{p_\lambda\}_{\lambda \in S}$ and $\{q_\lambda\}_{\lambda \in S}$ are the coefficients of p and q, respectively. For a set $S \subset \mathbb{Z}^d$ satisfying $S = -S$ we let $\mathbb{T}\mathrm{Pol}_S$ be the vector space over $\mathbb{R}$ consisting of all trigonometric polynomials p with Fourier support in S such that $p(z) \in \mathbb{R}$ for $z \in \mathbb{T}^d$. Equivalently, $p \in \mathbb{T}\mathrm{Pol}_S$ if $p(z) = \sum_{s \in S} p_s z^s$ satisfies $p_s = \overline{p_{-s}}$ for $s \in S$. With the inner product inherited from Pol_S the space $\mathbb{T}\mathrm{Pol}_S$ is a Hilbert space over $\mathbb{R}$. We will be especially interested in the case when S is of the form $S = \Lambda - \Lambda$, where $\Lambda \subset \mathbb{N}_0^d$.

Let us introduce the following cones:

$$\mathbb{T}\mathrm{Pol}_S^+ = \{p \in \mathbb{T}\mathrm{Pol}_S : p(z) \geq 0 \text{ for all } z \in \mathbb{T}^d\},$$

the cone consisting of all nonnegative valued trigonometric polynomials with support in S, and

$$\mathbb{T}\mathrm{Pol}_\Lambda^2 = \{p \in \mathbb{T}\mathrm{Pol}_S : p = \sum_{j=1}^r |q_j|^2, \text{ with } q_j \in \mathrm{Pol}_\Lambda\},$$

consisting of all trigonometric polynomials that are sums of squares of absolute values of polynomials with support in Λ. Obviously, $\mathbb{T}\mathrm{Pol}_\Lambda^2 \subseteq \mathbb{T}\mathrm{Pol}_{\Lambda-\Lambda}^+$.

Next we introduce two cones that are defined via Toeplitz matrices built from Fourier coefficients. In general, when we have a finite subset Λ of $\mathbb{Z}^d$ and a trigonometric polynomial $p(z) = \sum_{s\in S} p_s z^s$, where $S = \Lambda - \Lambda$, we can define the (multivariable) Toeplitz matrix

$$T_{p,\Lambda} = (p_{\lambda-\nu})_{\lambda,\nu\in\Lambda}.$$

It should be noticed that in this definition of the multivariable Toeplitz matrix there is some ambiguity in the order of rows and columns. We will only be interested in properties (primarily positive semidefiniteness) of the matrix that are not dependent on the order of the rows and columns, as long as the same order is used for both the rows and the columns. For example, consider $\Lambda = \{(0,0),(0,1),(1,0)\} \subseteq \mathbb{Z}^2$. If we order the elements of Λ as indicated, we get that

$$(p_{\lambda-\nu})_{\lambda,\nu\in\Lambda} = \begin{pmatrix} p_{0,0} & p_{0,-1} & p_{-1,0} \\ p_{0,1} & p_{0,0} & p_{-1,1} \\ p_{1,0} & p_{1,-1} & p_{0,0} \end{pmatrix}.$$

If we order Λ as $\{(0,1),(0,0),(1,0)\}$ we get the matrix

$$(p_{\lambda-\nu})_{\lambda,\nu\in\Lambda} = \begin{pmatrix} p_{0,0} & p_{0,1} & p_{-1,1} \\ p_{0,-1} & p_{0,0} & p_{-1,0} \\ p_{1,-1} & p_{1,0} & p_{0,0} \end{pmatrix},$$

which is of course the previous matrix with rows and columns 1 and 2 permuted. Later on, we may refer, for instance, to the $(0,0)$th row and the $(0,1)$th column, and even the $((0,0),(0,1))$th element of the matrix. In that terminology, we have that the (λ,ν)th element of the matrix $T_{p,\Lambda}$ is the number $p_{\lambda-\nu}$, explaining the convenience of this terminology. Of course, when $\Lambda = \{1,\ldots,n\}$ and the numbers 1 through n are ordered as usual, we get a classical Toeplitz matrix and the associated terminology is the usual one. In general, though, these "multivariable" Toeplitz matrices and the associated terminology take some getting used to. Even in the case when $\Lambda \subset \mathbb{Z}$ some care is needed in these definitions. For instance, for both subsets $\Lambda = \{0,1,3\}$ and $\Lambda' = \{0,1,2,3\}$ of $\mathbb{Z}$, we have that $S = \Lambda - \Lambda = \Lambda' - \Lambda' = \{-3,-2,-1,0,1,2,3\}$. For these sets we have that

$$T_{p,\Lambda} = \begin{pmatrix} p_0 & p_{-1} & p_{-3} \\ p_1 & p_0 & p_{-2} \\ p_3 & p_2 & p_0 \end{pmatrix}, \tag{1.3.1}$$

while

$$T_{p,\Lambda'} = \begin{pmatrix} p_0 & p_{-1} & p_{-2} & p_{-3} \\ p_1 & p_0 & p_{-1} & p_{-2} \\ p_2 & p_1 & p_0 & p_{-1} \\ p_3 & p_2 & p_1 & p_0 \end{pmatrix}. \tag{1.3.2}$$

The positive semidefiniteness of (1.3.1) does not guarantee positive semidefiniteness of (1.3.2). For example, when we take $p(z) = \frac{7}{10z^3} + \frac{7}{10z} + 1 + \frac{7z}{10} + \frac{7z^3}{10}$ it is easy to check that $T_{p,\Lambda}$ is positive semidefinite but $T_{p,\Lambda'}$ is not.

It should be noted that the condition $T_{p,\Lambda} \geq 0$ is equivalent to the statement that

$$\sum_{\lambda,\mu\in\Lambda} p_{\lambda-\mu} c_\mu \bar{c}_\lambda \geq 0 \tag{1.3.3}$$

for every complex sequence $\{c_\lambda\}_{\lambda\in\Lambda}$. Relation (1.3.3) transforms $T_{p,\Lambda} \geq 0$ into a condition that is independent of the ordering of Λ, and in the literature one may see the material addressed in this section presented in this way. We chose to use the presentation with the positive semidefinite Toeplitz matrices because it parallels the setting of positive semidefinite matrices as discussed in the previous section.

Next, we define the notion of extendability. Given $p(z) = \sum_{s\in S} p_s z^s$, we say that p is *extendable* if we can find p_m, $m \in \mathbb{Z}^d \setminus S$, such that for every finite set $J \subset \mathbb{Z}^d$ the Toeplitz matrix $(p_{\lambda-\nu})_{\lambda,\nu\in J}$ is positive semidefinite. Obviously, if $S = \Lambda - \Lambda$, then p being extendable implies that $T_{p,\Lambda} \geq 0$.

We are now ready to introduce the next two cones:

$$\mathcal{A}_\Lambda = \{p \in \mathbb{T}\mathrm{Pol}_{\Lambda-\Lambda} : T_{p,\Lambda} \geq 0\}$$

and

$$\mathcal{B}_S = \{p \in \mathbb{T}\mathrm{Pol}_S : p \text{ is extendable}\}.$$

Clearly, $\mathcal{B}_{\Lambda-\Lambda} \subseteq \mathcal{A}_\Lambda$. In general, though, we do not have equality. For instance,

$$p(z) = \frac{7}{10z^3} + \frac{7}{10z} + 1 + \frac{7z}{10} + \frac{7z^3}{10} \in \mathcal{A}_{\{0,1,3\}} \setminus \mathcal{B}_{\{-3,\ldots,3\}}. \tag{1.3.4}$$

The relationship between the four convex cones introduced above is the subject of the following result.

Proposition 1.3.1 *Let $\Lambda \subset \mathbb{Z}^d$ be finite and put $S = \Lambda - \Lambda$. Then the cones $\mathbb{T}\mathrm{Pol}^2_\Lambda$, $\mathbb{T}\mathrm{Pol}^+_S$, $\mathcal{B}_S$, and $\mathcal{A}_\Lambda$ are closed in $\mathbb{T}\mathrm{Pol}_S$, and we have*

$$\mathbb{T}\mathrm{Pol}^2_\Lambda \subseteq \mathbb{T}\mathrm{Pol}^+_S \subseteq \mathcal{B}_S \subseteq \mathcal{A}_\Lambda. \tag{1.3.5}$$

Proof. The closedness of $\mathbb{T}\mathrm{Pol}^+_S$, $\mathcal{B}_S$, and $\mathcal{A}_\Lambda$ is trivial. For the closedness of $\mathbb{T}\mathrm{Pol}^2_\Lambda$, observe first that $\dim(\mathbb{T}\mathrm{Pol}^+_S) = \mathrm{card} S =: m$. First we show that every element of $\mathbb{T}\mathrm{Pol}^2_\Lambda$ is a sum of at most m squares. Let $p \in \mathbb{T}\mathrm{Pol}^2_\Lambda$, and write

$$p = \sum_{j=1}^{r} |q_j|^2, \tag{1.3.6}$$

with each $q_j \in \mathbb{T}\mathrm{Pol}_\Lambda$. If $r \le m$ we are done, so assume that $r > m$. Since each $|q_j|^2$ is in $\mathbb{T}\mathrm{Pol}_S^+$, there exist real numbers r_j, not all 0, such that

$$\sum_{j=1}^{r} r_j |q_j|^2 = 0. \tag{1.3.7}$$

Without loss of generality we may assume that $|r_1| \ge |r_2| \ge \cdots \ge |r_m|$ and $|r_1| > 0$. Solving (1.3.7) for $|q_1|^2$ and substituting that into (1.3.6) we obtain

$$p = \sum_{j=2}^{r} \left(1 - \frac{r_j}{r_1}\right) |q_j|^2,$$

which is the sum of $r-1$ squares. Repeating these arguments we see that every $p \in \mathbb{T}\mathrm{Pol}_\Lambda^2$ is the sum of at most m squares.

If $p_n \in \mathbb{T}\mathrm{Pol}_\Lambda^2$ for $n \ge 1$ and p_n converges uniformly on $\mathbb{T}^d$ to p, there exist $q_{j,n} \in \mathrm{Pol}_\Lambda$ such that

$$p_n = \sum_{j=1}^{m} |q_{j,n}|^2, \quad n \ge 1. \tag{1.3.8}$$

The sequence p_n is uniformly bounded on Γ; hence so are $q_{j,n}$ by (1.3.8). Fix $k \in \Lambda$. It follows that there is a sequence $n_i \to \infty$ such that the kth Fourier coefficients of q_{j,n_i} converge to $g_{j,k}$, say, for $j = 1, \ldots, m$. Defining

$$\widetilde{q}_j(z) = \sum_{k \in \Lambda} g_{j,k} z^k$$

for $j = 1, \ldots, m$, we obviously have $p = \sum_{j=1}^m |\widetilde{q}_j|^2$. Thus $p \in \mathbb{T}\mathrm{Pol}_\Lambda^2$, and this proves $\mathbb{T}\mathrm{Pol}_\Lambda^2$ is closed.

The first and third inclusions in (1.3.5) are trivial. For the second inclusion, let $p \in \mathbb{T}\mathrm{Pol}_S^+$ and put $p_\nu = 0$ for $\nu \notin S$. With this choice we claim that $T_{p,J} \ge 0$ for all finite $J \subset \mathbb{Z}^d$. Indeed, for a complex sequence $v = (v_j)_{j \in J}$ define the polynomial $g(z) = \sum_{j \in J} v_j z^j$. Then

$$v^* T_{p,J} v = \frac{1}{(2\pi)^d} \int_{[0,2\pi]^d} p(e^{ij_1 t_1}, \ldots, e^{ij_d t_d}) |g(e^{ij_1 t_1}, \ldots, e^{ij_d t_d})|^2 dt \ge 0,$$

as $p(z) \ge 0$ for all $z \in \mathbb{T}^d$. This yields that p is extendable; thus $p \in \mathcal{B}_S$. $\square$

The next result identifies duality relations between the four cones.

Theorem 1.3.2 *With the earlier notation, the following duality relations hold in* $\mathbb{T}\mathrm{Pol}_S$*:*

(i) $(\mathbb{T}\mathrm{Pol}_\Lambda^2)^* = \mathcal{A}_\Lambda$*;*

(ii) $(\mathbb{T}\mathrm{Pol}_S^+)^* = \mathcal{B}_S$*.*

In order to prove this theorem we need a particular case of a result that is known as Bochner's theorem, which will be proved in its full generality as

Theorem 3.9.2. The following result is its particular version for $\mathbb{Z}^d$. For a finite Borel measure μ on $\mathbb{T}^d$, its *moments* are defined as

$$\widehat{\mu}(n) = \frac{1}{(2\pi)^d} \int\limits_{\mathbb{T}^d} z^{-n} d\mu(z), \ n \in \mathbb{Z}^d.$$

All measures appearing in this book are assumed to be regular.

Theorem 1.3.3 *Let μ be a finite positive Borel measure μ on $\mathbb{T}^d$. Then the Toeplitz matrices*

$$(\widehat{\mu}(j-k))_{j,k\in J}$$

are positive semidefinite for all finite sets $J \subset \mathbb{Z}^d$. Conversely, if $(c_j)_{j\in\mathbb{Z}^d}$ is a sequence of complex numbers such that for all finite $J \subset \mathbb{Z}^d$, the Toeplitz matrices

$$(c_{j-k})_{j,k\in J}$$

are positive semidefinite, then there exists a finite positive Borel measure μ on $\mathbb{T}^d$ such that $\widehat{\mu}(j) = c_j$, $j \in \mathbb{Z}^d$.

Proof. This result is a particular case of Theorem 3.9.2. A proof for this version may be found in [314]. □

We also need measures with finite support. For $\alpha \in \mathbb{T}^d$, let δ_α denote the *Dirac mass at* α (or *evaluation measure at* α). Thus for a Borel set $E \subseteq \mathbb{T}^d$ we have

$$\delta_\alpha(E) = \begin{cases} 1 & \text{if } \alpha \in E, \\ 0 & \text{otherwise.} \end{cases}$$

We will frequently use the fact that for a Borel measurable function $f : \mathbb{T}^d \to \mathbb{C}$, $\int_{\mathbb{T}^d} f d\delta_\alpha = f(\alpha)$. Applying this to the monomials $f(z) = z^{-n}$ we get that $\widehat{\delta}_\alpha(n) = \alpha^{-n}$ for every $\alpha \in \mathbb{T}^d$. A positive Borel measure on $\mathbb{T}^d$ is said to be supported on the finite subset $\{\alpha_1, \ldots, \alpha_n\} \subset \mathbb{T}^d$ if there exist positive constants $\rho_1, \ldots, \rho_n$ such that $\mu = \sum_{k=1}^n \rho_k \delta_{\alpha_k}$.

Proof of Theorem 1.3.2. It is clear that for $p \in \mathbb{T}\mathrm{Pol}_S$, we have that $p \in (\mathbb{T}\mathrm{Pol}_\Lambda^2)^*$ if and only if $\langle p, |q|^2\rangle \geq 0$ for all $q \in \mathrm{Pol}_\Lambda$. Let $q(z) = \sum_{\lambda\in\Lambda} c_\lambda z^\lambda$; then $\langle p, |q|^2\rangle = \sum_{\lambda,\mu\in\Lambda} p_{\lambda-\mu} c_\mu \overline{c}_\lambda = v^* T_{p,\Lambda} v$, where $v = (c_\lambda)_{\lambda\in\Lambda}$. Thus $\langle p, |q|^2\rangle \geq 0$ for all $q \in \mathrm{Pol}_\Lambda$ is equivalent to $T_{p,\Lambda} \geq 0$. But then it follows that $p \in (\mathbb{T}\mathrm{Pol}_\Lambda^2)^*$ if and only if $p \in \mathcal{A}_\Lambda$.

For proving (ii), let $p \in \mathcal{B}_S$, where $p(z) = \sum_{s\in S} p_s z^s$. As p is extendible, we can find p_ν, $\nu \in \mathbb{Z}^d \setminus S$ so that $(p_{j-k})_{j,k\in J} \geq 0$ for all finite $J \subset \mathbb{T}^d$. Apply Theorem 1.3.3 to obtain a finite positive Borel measure μ on $\mathbb{T}^d$ so that $\widehat{\mu}(j) = c_j$, $j \in \mathbb{Z}^d$. For every $q \in \mathbb{T}\mathrm{Pol}_S^+$, we have that

$$\langle q, p\rangle = \sum_{s\in S} q_s \widehat{\mu}(-s) = \int\limits_{\mathbb{T}^d} q(z) d\mu(z) \geq 0.$$

This yields $\mathbb{T}\mathrm{Pol}_S^+ \subseteq (\mathcal{B}_S)^*$. For the reverse inclusion, let $q \in (\mathcal{B}_S)^*$ and let $\alpha \in \mathbb{T}^d$. Put $p_{\alpha,S}(z) = \sum_{s\in S} \alpha^{-s} z^s$. Since $p_{\alpha,S} \in \mathcal{B}_S$, we have that

$0 \leq \langle q, p_{\alpha,S} \rangle = q(\alpha)$. This yields that $q(\alpha) \geq 0$ for all $\alpha \in \mathbb{T}^d$ and thus $q \in \mathbb{T}\mathrm{Pol}_S^+$. Hence we obtain $(\mathcal{B}_S)^* \subseteq \mathbb{T}\mathrm{Pol}_S^+$, which proves relation (ii). □

For one of the four cones it is easy to identify its extreme rays.

Theorem 1.3.4 *Let $\Lambda \subset \mathbb{Z}^d$ be a finite set and put $S = \Lambda - \Lambda$. The extreme rays of $\mathcal{B}_S$ are generated by the trigonometric polynomials $p_{\alpha,S}(z) = \sum_{s \in S} \alpha^{-s} z^s$, where $\alpha \in \mathbb{T}^d$.*

Proof. Let $\mathcal{C}_S$ be the closed cone generated by the trigonometric polynomials $p_{\alpha,S}$, $\alpha \in \mathbb{T}^d$. Clearly, $\mathcal{C}_S \subseteq \mathcal{B}_S$. Next, note that $q \in (\mathcal{C}_S)^*$ if only if $\langle q, p_{\alpha,S} \rangle = q(\alpha) \geq 0$ for all $\alpha \in \mathbb{T}^d$. But the latter is equivalent to the statement that $q \in \mathbb{T}\mathrm{Pol}_S^+$, and thus $(\mathcal{C}_S)^* = \mathbb{T}\mathrm{Pol}_S^+$. Taking duals we obtain from Theorem 1.3.2 that $\mathcal{C}_S = \mathcal{B}_S$. It remains to observe that each $p_{\alpha,S}$ generates an extreme ray of $\mathcal{B}_S$. This follows from the observation that the rank 1 positive semidefinite Toeplitz matrix $T_{p_{\alpha,S},\Lambda}$ cannot be written as a sum of positive semidefinite rank 1 matrices other than by using multiples of $T_{p_{\alpha,S},\Lambda}$ (see the proof of Proposition 1.1.4). □

As a consequence we obtain the following useful decomposition result.

Theorem 1.3.5 *Let $\Lambda \subset \mathbb{Z}^d$ be a finite set and put $S = \Lambda - \Lambda$. The trigonometric polynomial p belongs to $\mathcal{B}_S$ if and only if $T_{p,\Lambda}$ can be written as*

$$T_{p,\Lambda} = \sum_{j=1}^{r} \rho_j v_j v_j^*,$$

where $r \leq \mathrm{card} S$, $\rho_j > 0$ and $v_j = (\alpha_j^\lambda)_{\lambda \in \Lambda}$ with $\alpha_j \in \mathbb{T}^d$ for $j = 1, \ldots, r$.

Proof. Assume that $m > \mathrm{card} S$ and let $v = \sum_{j=1}^m \rho_j v_j v_j^*$, where $\rho_j > 0$ and $v_j = (\alpha_j^\lambda)_{\lambda \in \Lambda}$ with $\alpha_j \in \mathbb{T}^d$ for $j = 1, \ldots, m$. Since $m > \mathrm{card} S$, the matrices $v_j v_j^*$ are linearly dependent over $\mathbb{R}$. Without loss of generality, we may assume there exist $1 \leq l \leq m-1$ and nonnegative numbers $a_1, \ldots, a_l, b_{l+1}, \ldots, b_m$ such that

$$\sum_{i=1}^{l} a_i v_i v_i^* - \sum_{j=l+1}^{m} b_j v_j v_j^* = 0. \tag{1.3.9}$$

We order decreasingly $\frac{\rho_j}{b_j}$ for $b_j \neq 0$, and without loss of generality we assume that among them $\frac{\rho_m}{b_m}$ is the smallest. From (1.3.9) we obtain that

$$v_m v_m^* = \left(\sum_{i=1}^{l} a_i v_i v_i^* - \sum_{j=l+1}^{m-1} b_j v_j v_j^* \right) \frac{1}{b_m},$$

which implies that

$$\sum_{i=1}^{m} \rho_j v_j v_j^* = \sum_{i=1}^{l} \left(\rho_i + \frac{\rho_m a_i}{b_m} \right) v_i v_i^* + \sum_{j=l+1}^{m-1} \left(\rho_j - \frac{\rho_m b_j}{b_m} \right) v_j v_j^*.$$

Clearly all coefficients for $1 \leq i \leq l$ are positive. The minimality of $\frac{\rho_m}{b_m}$ implies that the coefficients for $l+1 \leq j \leq m-1$ are also positive, proving that every positive linear combination of matrices of the form $v_j v_j^*$ is a positive linear combination of at most card$S$ such matrices. Let $\mathcal{D}_S$ denote the cone of all matrices v of the form $\sum_{j=1}^r \rho_j v_j v_j^*$ with $\rho_j > 0$ and $r \leq$ card$S$. Using a similar argument as for proving the closedness of $\mathcal{B}_G$ in Proposition 1.2.1, one can show that $\mathcal{D}_S$ is closed. The details of this are left as Exercise 1.6.17. Using Theorem 1.3.4, we have that $p \in \mathcal{B}_S$ if and only if $T_{p,\Lambda} \in \mathcal{D}_S$, which is equivalent to the statement of theorem. □

1.3.1 The extension property

An important question is when equality occurs between the cones we have introduced. More specifically, for what $\Lambda \subseteq \mathbb{Z}^d$ do we have that $\mathbb{T}\mathrm{Pol}_\Lambda^2 = \mathbb{T}\mathrm{Pol}_{\Lambda-\Lambda}^+$ (or equivalently, $\mathcal{B}_{\Lambda-\Lambda} = \mathcal{A}_\Lambda$)? When these equalities occur, we say Λ has the *extension property*. It is immediate that this property is translation invariant, namely, that if Λ has the extension property then any set of the form $a + \Lambda$, $a \in \mathbb{Z}^d$, has it. Also, if $h : \mathbb{Z}^d \to \mathbb{Z}^d$ is a group isomorphism, then Λ has the extension property if and only if $h(\Lambda)$ does; we leave this as an exercise (see Exercise 1.6.30).

In case $\Lambda = \{0, 1, \ldots, n\} \subset \mathbb{Z}$, by the Fejér-Riesz factorization theorem (Theorem 1.1.5) we have that $\mathbb{T}\mathrm{Pol}_\Lambda^2 = \mathbb{T}\mathrm{Pol}_{\Lambda-\Lambda}^+$. This implies the following classical Carathéodory-Fejér theorem, which solves the so-called *truncated trigonometric moment problem.*

Theorem 1.3.6 *Let $c_j \in \mathbb{C}$, $j = -n, \ldots, n$, with $c_j = \overline{c_{-j}}$, $j = 0, \ldots, n$. Let T_n be the Toeplitz matrix $T_n = (c_{i-j})_{i,j=1}^n$. Then the following are equivalent.*

(i) There exists a finite positive Borel measure μ on $\mathbb{T}$ such that $c_k = \widehat{\mu}(k), k = -n, \ldots, n,$.

(ii) The Toeplitz matrix T_n is positive semidefinite.

(iii) The Toeplitz matrix T_n can be factored as $T_n = RDR^*$, where R is a Vandermonde matrix,

$$R = \begin{pmatrix} 1 & 1 & \ldots & 1 \\ \alpha_1 & \alpha_2 & \ldots & \alpha_r \\ \vdots & \vdots & & \vdots \\ \alpha_1^n & \alpha_2^n & \ldots & \alpha_r^n \end{pmatrix},$$

with $\alpha_j \in \mathbb{T}$ for $j = 1, \ldots, r$, and $\alpha_j \neq \alpha_p$ for $j \neq p$, and D is a $r \times r$ positive definite diagonal matrix.

(iv) there exist $\alpha_j \in \mathbb{T}$, $j = 1, \ldots, r$, with $\alpha_j \neq \alpha_p$, and $\rho_j > 0$, $j = 1, \ldots, r$, such that

$$c_l = \sum_{j=1}^r \rho_j \alpha_j^l, \quad |l| \leq n. \tag{1.3.10}$$

In case T_n is singular, $r = \operatorname{rank} T_n$, and the α_j are uniquely determined as the roots of the polynomial

$$p(z) = \det \begin{pmatrix} c_0 & \overline{c}_1 & \dots & \overline{c}_r \\ c_1 & c_0 & \dots & \overline{c}_{r-1} \\ \vdots & \vdots & \vdots & \vdots \\ c_{r-1} & c_{r-2} & \dots & \overline{c}_1 \\ 1 & z & \dots & z^r \end{pmatrix}.$$

In case T_n is nonsingular, one may choose $c_{n+1} = \overline{c}_{-n-1}$ such that $T_{n+1} = (c_{i-j})_{i,j=0}^{n+1}$ is positive semidefinite and singular, and apply the above to c_j, $|j| \leq n+1$. The measure μ can be chosen to equal $\mu = \sum_{j=1}^{r} \rho_j \delta_{\alpha_j}$, where ρ_j and α_j are as above.

Proof. Let $\Lambda = \{0, \dots, n\}$. By Theorem 1.1.5 we have that $\mathbb{T}\mathrm{Pol}_\Lambda^2 = \mathbb{T}\mathrm{Pol}_{\Lambda-\Lambda}^+$, and thus by duality $\mathcal{A}_\Lambda = \mathcal{B}_{\Lambda-\Lambda}$. Now (i) $\Rightarrow$ (ii) follows from Theorem 1.3.3. (ii) $\Rightarrow$ (iii) follows as $p \in \mathcal{A}_\Lambda$ implies $p \in \mathcal{B}_{\Lambda-\Lambda}$, which in turn implies (iii) by Theorem 1.3.5. The equivalence of (iii) and (iv) is immediate, which leaves the implication (iv) $\Rightarrow$ (i). The latter follows from Theorem 1.3.3. $\square$

In Exercise 1.6.20 we give an algorithm on how to find a solution to the truncated trigonometric moment problem that is a finite positive combination of Dirac masses.

We can also draw the following corollary, which will be useful in Section 5.7.

Corollary 1.3.7 *Let $z_1, \dots, z_n \in \mathbb{C} \setminus \{0\}$. Define $\sigma_{-n}, \dots, \sigma_n$ via*

$$\sigma_k = \sum_{j=1}^{n} z_j^k, \ k = -n, \dots, n.$$

Then $z_1, \dots, z_n$ are all distinct and lie on the unit circle $\mathbb{T}$ if and only if $\sigma = (\sigma_{i-j})_{i,j=0}^n \geq 0$ and rank $\sigma = n$.

Proof. First suppose that $z_1, \dots, z_n$ are all distinct and lie on the unit circle. Put

$$V = \begin{pmatrix} 1 & 1 & \dots & 1 \\ z_1 & z_2 & \dots & z_n \\ \vdots & \vdots & & \vdots \\ z_1^n & z_2^n & \dots & z_n^n \end{pmatrix}.$$

Then $\sigma = VV^* \geq 0$, and $\operatorname{rank} \sigma = \operatorname{rank} V = n$.

Next suppose that $\sigma \geq 0$ and $\operatorname{rank} \sigma = n$. By Theorem 1.3.6 we can write $\sigma = RDR^*$, with R and D as in Theorem 1.3.6 (where $r = n$). Put V as above, and

$$W = \begin{pmatrix} 1 & 1 & \dots & 1 \\ z_1^{-1} & z_2^{-1} & \dots & z_n^{-1} \\ \vdots & \vdots & & \vdots \\ z_1^{-n} & z_2^{-n} & \dots & z_n^{-n} \end{pmatrix}.$$

Then $\sigma = VW^T$. As $\operatorname{rank}\sigma = n$, we get that V and W must be of full rank. This yields that $z_1, \dots, z_n$ are different. Next, as $\sigma = VW = RDR^*$ has a one-dimensional kernel, there is a nonzero row vector $y = \begin{pmatrix} p_0 & \cdots & p_n \end{pmatrix}$ such that $y\sigma = 0$, and y is unique up to a multiplying with a nonzero scalar. But then we get that $yV = 0 = yR$, yielding that the n different numbers $z_1, \dots, z_n$ and the n different numbers $\alpha_1, \dots, \alpha_n$ are all roots of the nonzero polynomial $p(z) = \sum_{i=0}^n p_i z^i$. But then we must have that $\{z_1, \dots, z_n\} = \{\alpha_1, \dots, \alpha_n\} \subset \mathbb{T}$, finishing the proof. □

In the one variable case ($d = 1$), the finite subsets of $\mathbb{Z}$ which have the extension property are characterized by the following result.

Theorem 1.3.8 *Let Λ be a nonempty finite subset of $\mathbb{Z}$. Then Λ has the extension property if and only if $\Lambda = \{a, a+b, a+2b, \dots, a+kb\}$ for some $a, b, k \in \mathbb{Z}$ (namely, Λ is an arithmetic progression).*

Before we prove this result we will develop a useful technique to identify subsets of $\mathbb{Z}^d$ that do not have the extension property. For $\Lambda \subseteq \mathbb{Z}^d$ and $\emptyset \neq A \subset \mathbb{T}^d$, define

$$\Omega_\Lambda(A) = \{b \in \mathbb{T}^d \ : \ p(b) = 0 \text{ for all } p \in \operatorname{Pol}_\Lambda \text{ with } p|A \equiv 0\}, \tag{1.3.11}$$

where $p|A \equiv 0$ is short for $p(a) = 0$ for all $a \in A$. For $a \in \mathbb{T}^d$ and $\Lambda \subseteq \mathbb{Z}^d$, denote $L_a = \operatorname{row}(a^\lambda)_{\lambda\in\Lambda}$. Note that if $p(z) = \sum_{\lambda\in\Lambda} p_\lambda z^\lambda$, then $p(a) = L_a P$, where $P = \operatorname{col}(p_\lambda)_{\lambda\in\Lambda}$. Thus one may think of L_a as the linear functional on $\operatorname{Pol}_\Lambda$ that evaluates a polynomial at the point $a \in \mathbb{T}^d$. With this notation, we have

$$\Omega_\Lambda(A) = \{b \in \mathbb{T}^d \ : \ L_b P = 0 \text{ for all } P \in \operatorname{Pol}_\Lambda \text{ with } L_a P = 0 \ \text{ for all } a \in A\}.$$

The following is easy to verify.

Lemma 1.3.9 *For $\Lambda \subseteq \mathbb{Z}^d$ and $\emptyset \neq A, B \subset \mathbb{T}^d$, we have*

(i) $A \subseteq \Omega_\Lambda(A)$;

(ii) $A \subseteq B$ *implies* $\Omega_\Lambda(A) \subseteq \Omega_\Lambda(B)$;

(iii) $\Omega_\Lambda(A) = \Omega_{\Lambda+a}(A)$ *for all* $a \in \mathbb{Z}^d$.

For a nonempty set $S \subseteq \mathbb{Z}^d$ we let $G(S)$ denote the smallest (by inclusion) subgroup of $\mathbb{Z}^d$ containing S.

In case A is a singleton and $G(\Lambda - \Lambda) = \mathbb{Z}^d$, it is not hard to determine $\Omega_\Lambda(A)$.

Proposition 1.3.10 *Let $\Lambda \subseteq \mathbb{Z}^d$ be such that $G(\Lambda - \Lambda) = \mathbb{Z}^d$. Then for all $a \in \mathbb{Z}^d$ we have that $\Omega_\Lambda(\{a\}) = \{a\}$.*

Proof. Without loss of generality $0 \in \Lambda$ (use Lemma 1.3.9 (iii)). Let $b \in \Omega_\Lambda(\{a\})$, and let $0 \neq \lambda \in \Lambda$ (which must exist as $G(\Lambda - \Lambda) = \mathbb{Z}^d$). Introduce the polynomial $p(z) = z^\lambda - a^\lambda$. As $b \in \Omega_\Lambda(\{a\})$, we must have that $p(b) = 0$.

Thus we obtain that $a^\lambda = b^\lambda$ for all $\lambda \in \Lambda$. By taking products and inverses, and using that $G(\Lambda - \Lambda) = \mathbb{Z}^d$, we obtain that $a^\lambda = b^\lambda$ for all $\lambda \in \mathbb{Z}^d$. In particular, this holds for $\lambda = e_i$, $i = 1, \ldots, d$, where $e_i \in \mathbb{Z}^d$ has all zeros except for a 1 in the ith position. It now follows that $a_i = b_i$, $i = 1, \ldots, d$, and thus $a = b$. □

Exercise 1.6.31 shows that when $G(\Lambda - \Lambda) \neq \mathbb{Z}^d$ then $\Omega_\Lambda(\{a\})$ contains more than one element.

Theorem 1.3.11 *Let $\Lambda \subseteq \mathbb{Z}^d$ be a finite set such that $G(\Lambda - \Lambda) = \mathbb{Z}^d$. If Λ has the extension property, then for all finite subsets A of $\mathbb{T}^d$ either $\Omega_\Lambda(A) = A$ or* $\mathrm{card}\Omega_\Lambda(A) = \infty$.

Before starting the proof of Theorem 1.3.11, we need several auxiliary results. The range and kernel of a matrix T are denoted by $\mathrm{Ran}\, T$ and $\mathrm{Ker}\, T$, respectively.

Lemma 1.3.12 *Let $T \geq 0$, $0 \neq v \in \mathrm{Ran}\, T$. Let x be so that $Tx = v$. Then $\mu(v) := x^*Tx > 0$ is independent of the choice of x. Moreover, $T - \frac{1}{\mu(v)} vv^* \geq 0$ and $x \in \mathrm{Ker}\, (T - \frac{1}{\mu(v)} vv^*)$.*

Proof. Clearly, $x^*Tx \geq 0$. If $x^*Tx = 0$, we get that $\|T^{\frac{1}{2}}x\|^2 = x^*Tx = 0$ and thus $Tx = T^{\frac{1}{2}}T^{\frac{1}{2}}x = 0$, which is a contradiction. Thus $x^*Tx > 0$. Next, if $Tx = v = T\widetilde{x}$, then $x - \widetilde{x} \in \mathrm{Ker}\, T$, and thus

$$x^*Tx - \widetilde{x}^*T\widetilde{x} = (x - \widetilde{x})Tx + \widetilde{x}^*T(x - \widetilde{x}) = 0.$$

Next,

$$\begin{pmatrix} T & v \\ v^* & \mu(v) \end{pmatrix} = \begin{pmatrix} I \\ x^* \end{pmatrix} T \begin{pmatrix} I & x \end{pmatrix} \geq 0,$$

and thus, by Lemma 1.2.5, $T - \frac{vv^*}{\mu(v)} \geq 0$.

Finally,

$$x^* \left(T - \frac{vv^*}{\mu(v)} \right) x = x^*Tx - \frac{(x^*Tx)^2}{\mu(v)} = 0,$$

and since $T - \frac{vv^*}{\mu(v)} \geq 0$ it follows that $x \in \mathrm{Ker}\, (T - \frac{vv^*}{\mu(v)})$. □

Lemma 1.3.13 *Let*

$$T = (c_{\lambda-\mu})_{\lambda,\mu\in\Lambda} \geq 0,$$

and

$$\Sigma_T = \{a \in \mathbb{T}^d \ : L_a P = 0 \text{ for all } P \in \mathrm{Ker}\, T\}.$$

Then $a \in \Sigma_T$ if and only if $L_a^* \in \mathrm{Ran}\, T$.

Proof. As $T = T^*$, we have $\mathrm{Ran}\, T = (\mathrm{Ker}\, T)^\perp$, and the lemma follows. □

The extension property implies that $\Sigma_T \neq \emptyset$ and has a strong consequence regarding the possible forms of the decompositions as in Theorem 1.3.5.

Proposition 1.3.14 *Let $\Lambda \subseteq \mathbb{Z}^d$ be a finite set that has the extension property, and let*

$$0 \neq T = (c_{\lambda-\mu})_{\lambda,\mu\in\Lambda} \geq 0.$$

Put $m = \operatorname{rank} T$, and let Σ_T as in Lemma 1.3.13. Then $\Sigma_T \neq \emptyset$, and for any $b_1 \in \Sigma_T$ there exist $b_2, \ldots, b_m \in \Sigma_T$ and $\mu_1, \ldots, \mu_m > 0$ such that

$$T = \sum_{k=1}^{m} \frac{1}{\mu_k} L_{b_k}^* L_{b_k}.$$

Note that necessarily $L_{b_1}, \ldots L_{b_m}$ are linearly independent.

Proof. We first show that $\Sigma_T \neq \emptyset$. As Λ has the extension property, we have that $\mathcal{B}_{\Lambda-\Lambda} = \mathcal{A}_\Lambda$. As $T \in \mathcal{A}_\Lambda = \mathcal{B}_{\Lambda-\Lambda}$, we obtain from Theorem 1.3.5 that we may represent T as

$$T = \sum_{k=1}^{r} \rho_k L_{a_k}^* L_{a_k},$$

where $a_1, \ldots, a_r \in \mathbb{T}^d$ and $\rho_1, \ldots, \rho_r > 0$. Now, if $P \in \operatorname{Ker} T$, we get that $P^*TP = \sum_{k=1}^{r} \rho_k |L_{a_k} P|^2 = 0$ and thus $L_{a_k} P = 0$, $k = 1, \ldots, r$. This yields that $a_1, \ldots, a_r \in \Sigma_T$.

Let now $b_1 \in \Sigma_T$. By Lemma 1.3.13, $0 \neq L_{b_1}^* = Tx$ for some $x \neq 0$. If we put $\mu_1 = x^*Tx$, we have by Lemma 1.3.12 that $\mu_1 > 0$, $\widetilde{T} := T - \frac{1}{\mu_1} L_{b_1}^* L_{b_1} \geq 0$, and $x \in \operatorname{Ker} \widetilde{T}$. In particular, since $Tx = L_{b_1} \neq 0$ and $\operatorname{Ker} T \subseteq \operatorname{Ker} \widetilde{T}$ (as $0 \leq \widetilde{T} \leq T$), we get that $\dim \operatorname{Ker} \widetilde{T} = \dim \operatorname{Ker} T + 1$, and thus $\operatorname{rank} \widetilde{T} = \operatorname{rank} T - 1$.

If $\widetilde{T} = 0$, we are done. If not, note that $\Sigma_{\widetilde{T}} \subseteq \Sigma_T$, and repeat the above with $\widetilde{T}$ instead of T. As in each step the rank decreases by 1, we are done in at most m steps. □

Lemma 1.3.15 *Let $x = (x_1, \ldots, x_m)^T \in \mathbb{C}^m$, and suppose x has all nonzero components. Then there is no nonempty finite subset $F \subseteq \mathbb{C}^m$ with the properties that $0 \notin F$ and that for every $d_1, \ldots, d_m > 0$ there exists $y = (y_1, \ldots, y_m)^T \in F$ satisfying $\sum_{i=1}^{m} d_i x_i \overline{y}_i = 0$.*

Proof. We prove the claim by induction. When $m = 1$ the statement is obviously true. Suppose now that the result holds for m, and we prove it for $m+1$. Let $x = (x_1, \ldots, x_{m+1})^T \in \mathbb{C}^{m+1}$ be given with $x_i \neq 0$, $i = 1, \ldots, m+1$, and suppose $F \subset \mathbb{C}^{m+1}$ is finite. We can assume $(y_1, \ldots, y_m) \neq (0, \ldots, 0)$ for any $(y_1, \ldots, y_m, y_{m+1})^T \in F$. Indeed, otherwise we would have that $d_{m+1} x_{m+1} \overline{y}_{m+1} \neq 0$ for all $d_{m+1} > 0$, which is impossible since $x_{m+1} \neq 0$ and $y_{m+1} \neq 0$ (as $0 \notin F$). But then we can apply our induction hypothesis, and conclude that there exist $d_1^0, \ldots, d_m^0 > 0$ such that

$$\sum_{i=1}^{m} d_i^0 x_i \overline{y}_i \neq 0$$

for all $(y_1, \ldots, y_m, y_{m+1})^T \in F$. If the equation

$$\sum_{i=1}^{m} d_i^0 x_i \overline{y}_i + d_{m+1} x_{m+1} \overline{y}_{m+1} = 0$$

holds, then necessarily $y_{m+1} \neq 0$ and

$$d_{m+1} = -\frac{\sum_{i=1}^{m} d_i^0 x_i \overline{y}_i}{x_{m+1} \overline{y}_{m+1}}.$$

Since the right-hand side can take on only a finite number of values, we get that there must exists a $d_{m+1}^0 > 0$ so that

$$\sum_{i=1}^{m+1} d_i^0 x_i \overline{y}_i \neq 0$$

for all $(y_1, \ldots, y_m, y_{m+1})^T \in F$, proving our claim. □

We are now ready to prove Theorem 1.3.11.

Proof of Theorem 1.3.11. We do this by induction on $m = \text{card} A$. When $m = 1$ it follows from Proposition 1.3.10. Next suppose that the result has been proven for sets up to cardinality $m - 1$.

Let $A = \{a_1, \ldots, a_m\}$. If $L_{a_1}, \ldots, L_{a_m}$ are linearly dependent, then, without loss of generality, L_{a_m} belongs to the span of $L_{a_1}, \ldots, L_{a_{m-1}}$. But then, for $p(z) = \sum_{\lambda \in \Lambda} p_\lambda z^\lambda \in \text{Pol}_\Lambda$ and $P = \text{col}(p_\lambda)_{\lambda \in \Lambda}$, we have that $p(a_k) = L_{a_k} P = 0$, $k = 1, \ldots, m-1$, implies that $p(a_m) = L_{a_m} P = 0$. This gives that $\Omega_\Lambda(\{a_1, \ldots, a_m\}) = \Omega_\Lambda(\{a_1, \ldots, a_{m-1}\})$, and one can finish this case by using the induction assumption.

Next assume that $L_{a_1}, \ldots, L_{a_m}$ are linearly independent, and that the set $\Omega_\Lambda(\{a_1, \ldots, a_m\})$ is finite (otherwise we are done). By Lemma 1.3.9(iii) this implies that $\Omega_\Lambda(\{a_1, \ldots, a_{m-1}\})$ is a finite set as well. As $L_{a_1}, \ldots, L_{a_m}$ are linearly independent, we can find $U_1, \ldots, U_m$ so that

$$\begin{pmatrix} L_{a_1} \\ \vdots \\ L_{a_m} \end{pmatrix} \begin{pmatrix} U_1 & \cdots & U_m \end{pmatrix} = I_m,$$

or, equivalently, $L_{a_i} U_j = \delta_{ij}$ with δ_{ij} being the Kronecker delta. Let $b_1 \in \Omega_\Lambda(\{a_1, \ldots, a_m\})$. We claim that $L_{b_1} \in \text{Span}\{L_{a_1}, \ldots, L_{a_m}\}$. It suffices to show that $L_{b_1} P = 0$ for all P with $L_{a_1} P = \cdots = L_{a_m} P = 0$. But this follows directly from the definition of $\Omega_\Lambda(\{a_1, \ldots, a_m\})$.

Next, suppose that $L_{b_1} U_i = 0$ for some $i = 1, \ldots, m$. Without loss of generality, $L_{b_1} U_m = 0$. As L_{b_1} is in the span of $L_{a_1}, \ldots, L_{a_m}$, we may write $L_{b_1} = \sum_{i=1}^{m} z_i L_{a_i}$ for some complex numbers z_i. But then $0 = L_{b_1} U_m = \sum_{i=1}^{m} z_i L_{a_i} U_m = z_m$. Thus it follows that L_{b_1} lies in the span of $L_{a_1}, \ldots, L_{a_{m-1}}$. This, as before, gives that $b_1 \in \Omega_\Lambda(\{a_1, \ldots, a_{m-1}\})$. As $\text{card}\Omega_\Lambda(\{a_1, \ldots, a_{m-1}\}) < \infty$ we get by the induction assumption that $b_1 \in \Omega_\Lambda(\{a_1, \ldots, a_{m-1}\}) = \{a_1, \ldots, a_{m-1}\}$, and we are done.

Finally, we are in the case when $L_{b_1}U_i \neq 0$ for all $i = 1, \dots, m$. We will show that this case cannot occur as we will reach a contradiction. Let $D = \mathrm{diag}(d_i)_{i=1}^m > 0$, and consider

$$T(D) = \sum_{i=1}^m d_i L_{a_i}^* L_{a_i}.$$

Note that $\Sigma_{T(D)} = \Omega_\Lambda(\{a_1, \dots, a_m\})$, where Σ_T is defined in Proposition 1.3.14. By Proposition 1.3.14 there exist $\mu_k = \mu_k(D) > 0$, $k = 1, \dots, m$, and $b_2(D), \dots, b_m(D) \in \Omega_\Lambda(\{a_1, \dots, a_m\})$ such that

$$T(D) = \frac{1}{\mu_1(D)} L_{b_1}^* L_{b_1} + \sum_{k=2}^m \frac{1}{\mu_k(D)} L_{b_k(D)}^* L_{b_k(D)}.$$

Thus

$$\sum_{i=1}^m d_i L_{a_i}^* L_{a_i} = \frac{1}{\mu_1(D)} L_{b_1}^* L_{b_1} + \sum_{k=2}^m \frac{1}{\mu_k(D)} L_{b_k(D)}^* L_{b_k(D)}.$$

Multiplying this equation with $U = \begin{pmatrix} U_1 & \cdots & U_m \end{pmatrix}$ on the right and with U^* on the left, we get that

$$D = Q^* \mathrm{diag}(\mu_i(D))_{i=1}^m Q,$$

where

$$Q = \begin{pmatrix} L_{b_1} \\ L_{b_2(D)} \\ \vdots \\ L_{b_m(D)} \end{pmatrix} U.$$

Thus

$$I_m = (Q^* \mathrm{diag}(\mu_i(D))_{i=1}^m)(QD^{-1}),$$

and as all matrices in this equation are $m \times m$ we also get

$$I_m = (QD^{-1})(Q^* \mathrm{diag}(\mu_i(D))_{i=1}^m).$$

Looking at the $(1, 2)$ entry of this equality (remember $m \geq 2$) and dividing by $\mu_2(D) > 0$, we get that for all for all $D = \mathrm{diag}(d_i)_{i=1}^m > 0$ there exists $b_2(D) \in \Omega_\Lambda(\{a_1, \dots, a_m\})$ such that

$$\begin{pmatrix} \frac{1}{d_1} L_{b_1} U_1 & \cdots & \frac{1}{d_m} L_{b_1} U_m \end{pmatrix} U^* L_{b_2(D)}^* = 0.$$

As $L_{b_1}U_i \neq 0$, $i = 1, \dots, m$, we get by Lemma 1.3.15 that there must be infinitely many different vectors $U^* L_{b_2(D)}$. But as each $b_2(D)$ lies in the finite set $\Omega_\Lambda(\{a_1, \dots, a_m\})$, we have a contradiction. □

Corollary 1.3.16 *Assume $\Lambda \subset \mathbb{Z}^d$ is a finite subset such that $G(\Lambda - \Lambda) = \mathbb{Z}^d$. If there exist k linearly independent polynomials $p_1, \dots, p_k \in \mathrm{Pol}_\Lambda$ such that the set $A = \cap_{i=1}^k \{a \in \mathbb{T}^d : p_i(a) = 0\}$ is a finite set with cardinality greater than* $\mathrm{card}\Lambda - k$, *then Λ fails to have the extension property.*

Proof. The card$A \times$ cardΛ matrix $\mathrm{col}(L_a)_{a\in A}$ has a kernel of dimension at least k, and thus rank $\mathrm{col}(L_a)_{a\in A} \leq \mathrm{card}\Lambda - k < \mathrm{card}A$. Thus the set of linear functionals $\{L_a\}_{a\in A}$ are linearly dependent. Let $\{a_1, \ldots, a_m\}$ be a maximal subset of A for which the set $\{L_{a_i}\}_{i=1}^m$ is linearly independent. Then $A = \Omega_\Lambda(\{a_1, \ldots, a_m\})$ is a finite set of cardinality greater than card$A$. Thus by Theorem 1.3.11, Λ does not have the extension property. □

We now have the techniques to easily prove Theorem 1.3.8.

Proof of Theorem 1.3.8. Since having the extension property is a translation invariant property, we may assume that $a = 0$. When $b = 1$, $\Lambda = \{0, 1, \ldots, k\}$ has the extension property by Theorem 1.3.6. Next replacing z by z^b, it easily follows that $\Lambda = \{0, b, 2b, \ldots, kb\}$ has the extension property.

Assume now that $\Lambda \subset \mathbb{Z}$ has the extension property, $G(\Lambda - \Lambda) = \mathbb{Z}$, 0 is the smallest element of Λ, and N is the largest. Then $z^N - 1 \in \mathrm{Pol}_\Lambda$ has N roots on $\mathbb{T}$, and consequently Corollary 1.3.16 implies that Λ has at least $N + 1$ elements. Thus $\Lambda = \{0, 1, \ldots, N\}$. If $G(\Lambda - \Lambda) \neq \mathbb{Z}$, then, letting b be the greatest common divisor of the nonzero elements of Λ, it is easy to see that that $G(\Lambda - \Lambda) = b\mathbb{Z}$. By considering polynomials in z^b as opposed to polynomials in z, one easily reduces it to the above case and sees that Λ must equal $\{0, b, 2b, \ldots, kb\}$, where $k = \frac{N}{b}$. □

In the multivariable case we consider finite sets Λ of the form

$$R(N_1, \ldots, N_d) := \{0, \ldots, N_1\} \times \cdots \times \{0, \ldots, N_d\}, \tag{1.3.12}$$

where R stands for rectangular.

Now that the one-variable case has been settled, let us first consider a three-variable case.

Theorem 1.3.17 *The extension property fails for $\Lambda = R(N_1, N_2, N_3)$, for every $N_1, N_2, N_3 \geq 1$.*

Proof. The set of the common zeros of the polynomials $(z_1 - 1)(z_2z_3 + 1)$, $(z_2 - 1)(z_1z_2 + 1)$, and $(z_3 - 1)(z_1z_2 + 1)$ consist of the points in $\mathbb{T}^3$:

$$\alpha_1 = (1, 1, 1),\ \alpha_2 = (1, -1, -1),\ \alpha_3 = (-1, 1, -1),$$

$$\alpha_4 = (-1, -1, 1),\ \alpha_5 = (i, i, i),\ \alpha_6 = (-i, -i, -i).$$

This implies that the set of common zeros in $\mathbb{T}^3$ of the polynomials $(z_1^{N_1} - 1)(z_2^{N_2}z_3^{N_3} + 1)$, $(z_2^{N_2} - 1)(z_1^{N_1}z_3^{N_3} + 1)$, and $(z_3^{N_3} - 1)(z_1^{N_1}z_2^{N_2} - 1)$ belonging to Pol_Λ has cardinality $6N_1N_2N_3$. Since card$\Lambda = (N_1 + 1)(N_2 + 1)(N_3 + 1)$, Corollary 1.3.16 with $k = 3$ that the extension property fails for $R(N_1, N_2, N_3)$ whenever

$$6N_1N_2N_3 > (N_1 + 1)(N_2 + 1)(N_3 + 1) - 3,$$

an inequality that is true when $N_i \geq 1$, $i = 1, 2, 3$. □

The remainder of this section is devoted to proving the following characterization of all the finite subsets of $\mathbb{Z}^2$ that have the extension property. We say that two finite subsets Λ_1 and Λ_2 of $\mathbb{Z}^2$ are *isomorphic* if there exists a group isomorphism Φ on $\mathbb{Z}^2$ so that $\Phi(\Lambda_1) = \Lambda_2$.

Theorem 1.3.18 *A finite set $\Lambda \subseteq \mathbb{Z}^2$ with $G(\Lambda - \Lambda) = \mathbb{Z}^2$ has the extension property if and only if for some $a \in \mathbb{Z}^2$ the set $\Lambda - a$ is isomorphic to $R(0,n)$, $R(1,n)$, or $R(1,n) \setminus \{(1,n)\}$, for some $n \in \mathbb{N}$.*

We prove this result in several steps. We use the following terminology. Let $R \subset \mathbb{Z}^d$. We say that $(c_k)_{k \in R-R}$ is a *positive semidefinite sequence* with respect to R if $(c_{k-l})_{k,l \in L} \geq 0$ for all finite subsets L of R. Clearly, if R is finite it suffices to check whether $(c_{k-l})_{k,l \in R} \geq 0$. Often the choice of the set R is clear, in which case we will just say that $(c_k)_{k \in R-R}$ is a positive semidefinite sequence.

Theorem 1.3.19 *For every $n \geq 1$, $R(1,n)$ has the extension property.*

Proof. Let $S = R(1,n) - R(1,n)$ and let $(c_{kl})_{(k,l) \in S}$ be a positive semidefinite sequence. Thus

$$\widetilde{C}_n = \begin{pmatrix} C_0 & C_1^* & \cdots & C_{n-1}^* & C_n^* \\ C_1 & C_0 & \ddots & & C_{n-1}^* \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ C_{n-1} & & \ddots & \ddots & C_1^* \\ C_n & C_{n-1} & \cdots & C_1 & C_0 \end{pmatrix} \geq 0, \tag{1.3.13}$$

where

$$C_j = \begin{pmatrix} c_{0j} & c_{1j} \\ c_{-1,j} & c_{0j} \end{pmatrix}, \ 0 \leq j \leq n.$$

Consider now the partial matrix

$$\begin{pmatrix} c_{00} & P^*C_1^* & \cdots & P^*C_n^* & X^* \\ C_1P & C_0 & \cdots & C_{n-1}^* & C_n^* \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ C_nP & C_{n-1} & \cdots & C_0 & C_1^* \\ X & C_n & \cdots & C_1 & C_0 \end{pmatrix}, \tag{1.3.14}$$

where $P = \left(\begin{smallmatrix} 0 \\ 1 \end{smallmatrix}\right)$ and X is the unknown. Since $\widetilde{C}_n$ is positive semidefinite, the partial matrix (1.3.14) is partially positive semidefinite. By Theorem 1.2.10 (i) $\Rightarrow$ (ii) we can find an $X = \left(\begin{smallmatrix} \beta \\ \alpha \end{smallmatrix}\right)$ so that (1.3.14) is positive semidefinite. Using the Toeplitz structure of C_j, $0 \leq j \leq n$, it is now not hard to see that the following matrix is positive semidefinite as well:

$$\begin{pmatrix} C_0 & C_1^* & \cdots & C_n^* & Y^* \\ C_1 & C_0 & \cdots & C_{n-1}^* & C_n^*Q \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ C_n & C_{n-1} & \cdots & C_0 & C_1^*Q \\ Y & Q^*C_n & \cdots & Q^*C_1 & c_{00} \end{pmatrix}, \tag{1.3.15}$$

where $Q = \left(\begin{smallmatrix} 1 \\ 0 \end{smallmatrix}\right)$ and $Y = \begin{pmatrix} \alpha & \beta \end{pmatrix}$. Indeed, (1.3.15) may be obtained from (1.3.14) by reversing the rows and columns and taking complex conjugates

entrywise. Next consider

$$\widetilde{C}_{n+1} = \begin{pmatrix} C_0 & C_1^* & \cdots & C_n^* & C_{n+1}^* \\ C_1 & C_0 & \cdots & C_{n-1}^* & C_n^* \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ C_n & C_{n-1} & \cdots & C_0 & C_1^* \\ C_{n+1} & C_n & \cdots & C_1 & C_0 \end{pmatrix},$$

where

$$C_{n+1} = \begin{pmatrix} \alpha & \beta \\ \gamma & \alpha \end{pmatrix}$$

with γ as an unknown. As (1.3.14) and (1.3.15) are positive semidefinite, we have that $\widetilde{C}_{n+1}$ is partially positive semidefinite. Applying again Theorem 1.2.10 (i) $\Rightarrow$ (ii) we can find a γ so that $\widetilde{C}_{n+1}$ is positive semidefinite. Define now $c_{0,n+1} = \alpha$, $c_{1,n+1} = \beta$, and $c_{-1,n+1} = \gamma$. This way we have extended the given positive sequence to a positive sequence supported in the set $R(1, n+1) - R(1, n+1)$. We repeat successively the previous construction for $n+1, n+2, \ldots$. This process produces a positive semidefinite extension of the given sequence supported in the infinite band $I_1 = R_1 - R_1$, where $R_1 = \{0, 1\} \times \mathbb{N}$. Thus, by Exercise 1.6.34, $R(1, n)$ has the extension property. This completes the proof. $\square$

In Exercise 2.9.33 we will point out an alternative proof for Theorem 1.3.19 that will make use of the Fejér-Riesz factorization theorem (Theorem 2.4.16).

Our next result introduces a new class of sets with the extension property.

Theorem 1.3.20 *For every $n \geq 1$, $R(1, n) \setminus \{(1, n)\}$ has the extension property.*

Proof. Let $(c_{kl})_{(k,l) \in R(1,n) \setminus \{(1,n\} - R(1,n) \setminus \{(1,n)\}}$ be an arbitrary positive semidefinite sequence. By Theorem 1.3.19, it is sufficient to prove that the sequence can be extended to a positive semidefinite sequence supported in $S = R(1, n) - R(1, n)$. The Toeplitz form associated with this sequence is

$$\widetilde{D}_n = \begin{pmatrix} C_0 & C_1^* & \cdots & C_{n-1}^* & D_n^* \\ C_1 & C_0 & \cdots & C_{n-2}^* & D_{n-1}^* \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ C_{n-1} & C_{n-2} & \cdots & C_0 & D_1^* \\ D_n & D_{n-1} & \cdots & D_1 & c_{00} \end{pmatrix},$$

where

$$C_j = \begin{pmatrix} c_{0j} & c_{1j} \\ c_{-1,j} & c_{0j} \end{pmatrix}$$

and $D_j = \begin{pmatrix} c_{0j} & c_{1j} \end{pmatrix}$.

It is easy to observe that deleting the first row and first column (resp., the last row and last column) of the matrix (1.3.13) of a sequence supported

in $R(1,n) - R(1,n)$, one obtains (a permuted version of) the matrix $\widetilde{D}_n$. Thus every positive semidefinite matrix $\widetilde{D}_n$ can be extended to a positive semidefinite matrix of the form (1.3.13), and this completes the proof. □

We would like to make a remark at this point.

Remark 1.3.21 Theorems 1.3.19 and 1.3.20 are true for scalar matrices only. They are not true when the entries are 2×2 or larger. Take for instance in Theorem 1.3.20, $\Lambda = \{(0,0),(0,1),(1,0)\}$, $C_{00} = \left(\begin{smallmatrix} 1 & 0 \\ 0 & 1 \end{smallmatrix}\right)$, and C_{01} and C_{10} to be two noncommuting unitary matrices. Then, in order to be positive semidefinite with respect to Λ, by Lemma 1.2.5, a sequence must satisfy $C_{-1,1} = C_{01}^* C_{10}$. We will show that no C_{11} exists for which the matrix of the sequence restricted to $\Lambda' = \{(0,0),(0,1),(1,0),(1,1)\}$ is positive semidefinite. Indeed, by a Schur complement type argument, we must have that C_{11} is unitary and $C_{11} = C_{01}C_{10} = C_{10}C_{01}$. The last equality does not hold, because C_{01} and C_{10} do not commute by assumption.

Theorem 1.3.22 *The set $R(N_1, N_2)$ has the extension property if and only if* $\min(N_1, N_2) = 1$.

We first need an auxiliary result.

Lemma 1.3.23 *The real polynomial*

$$F(s,t) = s^2t^2(s^2 + t^2 - 1) + 1$$

is strictly positive but it is not the sum of squares of polynomials with real coefficients.

Proof. It is an easy exercise (see Exercise 1.6.25) to show that F takes on the minimum value $\frac{26}{27}$ at $s = \pm\frac{1}{\sqrt{3}}$ and $t = \pm\frac{1}{\sqrt{3}}$.

Assume that

$$F = F_1^2 + \cdots + F_n^2 \tag{1.3.16}$$

for some real polynomials $F_1,\ldots,F_n$. From $F(s,0) = F(0,t) = 1$ it follows that $F_i(s,0)$ and $F_i(0,t)$ are constant for $i = 1,\ldots,n$, so

$$F_i(s,t) = a_i + stH_i(s,t) \tag{1.3.17}$$

for some constant a_i and some polynomial H_i of degree at most one. Substituting (1.3.17) into (1.3.16) and comparing the coefficients, we obtain

$$s^2t^2(s^2 + t^2 - 1) = s^2t^2 \sum_{k=1}^{n} H_i^2(s,t).$$

This is a contradiction as the right side is always nonnegative while the left is negative if $st \neq 0$ and $s^2 + t^2 < 1$. □

Proof of Theorem 1.3.22. We prove the theorem by showing that for $\Lambda = R(N_1, N_2)$ and $S = \Lambda - \Lambda$, $\mathbb{T}\mathrm{Pol}_S^+ \neq \mathbb{T}\mathrm{Pol}_\Lambda$.

Let X be the linear space of all real polynomials of the form $F(s,t) = \sum_{m=0}^{2N_1}\sum_{n=0}^{2N_2} a_{mn}s^m t^n$. Every $p \in \mathbb{T}\mathrm{Pol}_S$ is of the form

$$p(z,w) = \sum_{m=-N_1}^{N_1}\sum_{n=-N_2}^{N_2} c_{mn}z^m w^n$$

with $c_{-m,-n} = \overline{c}_{mn}$. Define the linear map $\Psi : \mathbb{T}\mathrm{Pol}_S \to X$ by

$$(\Psi(f))(s,t) = (1+s^2)^{N_1}(1+t^2)^{N_2} f\left(\frac{s+i}{s-i}, \frac{t+i}{t-i}\right).$$

It is clear Ψ is one-to-one, and since $\dim_{\mathbb{R}}(\mathbb{T}\mathrm{Pol}_S) = \dim_{\mathbb{R}} X = (2N_1 + 1)(2N_2+1)$, it follows that Ψ is a linear isomorphism of $\mathbb{T}\mathrm{Pol}_S$ onto X.

Assume that $f \in \mathbb{T}\mathrm{Pol}_\Lambda^2$. Then $f = \sum_{j=1}^r |g_j|^2$, where each $g_j \in \mathrm{Pol}_\Lambda$. Define for $j = 1, \ldots, r$,

$$G_j(s,t) = (s-i)^{N_1}(t-i)^{N_2} g_j\left(\frac{s+i}{s-i}, \frac{t+i}{t-i}\right).$$

Each G_j is a complex polynomial of degree at most N_1 in s and at most N_2 in t, and

$$(\Psi(f))(s,t) = \sum_{j=1}^r |G_j(s,t)|^2.$$

Define $F = \Psi(f)$ and let $G_j = u_j + iv_j$, where u_j and v_j are polynomials with real coefficients. We have then $F \in X$, and $F = \sum_{j=1}^r (u_j^2 + v_j^2)$, so F is a sum of squares. This cannot happen when F is the polynomial in Lemma 1.3.23, and this implies $\mathbb{T}\mathrm{Pol}_S^+ \neq \mathbb{T}\mathrm{Pol}_\Lambda^2$. □

Theorem 1.3.24 *Let $\Lambda \subset \mathbb{Z}^2$ be a finite subset containing the points $(0,0)$, $(m,0)$, $(0,n)$, and (m,n), where $m,n \geq 2$. If $\mathrm{card}\Lambda \leq (m+1)(n+1)$ and $G(S) = \mathbb{Z}^2$, then the extension property fails for Λ. In particular, the extension property fails for $R(m,n)$ for $m,n \geq 2$.*

Proof. The polynomial $1 + z_1 + z_2$ has two roots on $\mathbb{T}^2$, $(\alpha, \overline{\alpha})$ and $(\overline{\alpha}, \alpha)$, where $\alpha = e^{\frac{2\pi i}{3}}$. Therefore, the polynomial $1 + z_1^m + z_2^n \in \mathrm{Pol}_\Lambda$ has $2mn$ roots on $\mathbb{T}^2$. If $m,n \geq 2$ and either $m \geq 3$ or $n \geq 3$, we have that $\mathrm{card}(\Lambda) \leq (m+1)(n+1) \leq 2mn$, and so Λ fails the extension property by Corollary 1.3.16. When $m = n = 2$, the condition in Corollary 1.3.16 with $k = 2$ is verified by the polynomials $1 - z_1^2 z_2^2$ and $z_1^2 - z_2^2$, which have 8 common zeros on $\mathbb{T}^2$, proving $R(2,2)$ also fails the extension property. □

A *partial Toeplitz matrix* is a partial matrix that is Toeplitz at the extent to which it is specified; that is, all specified entries lie along certain *specified diagonals*, and all the entries along a specified diagonal have the same value. Positive semidefinite Toeplitz completions are of special interest. The trigonometric moment problem is equivalent to the fact that every partially positive Toeplitz matrix with entries specified in a band $|i-j| \leq k$, for some specified k, admits a positive Toeplitz completion.

By the *pattern* of a partial Toeplitz matrix we mean the set of its specified diagonals. The main diagonal is always assumed to be specified. Thus the pattern of a partially positive semidefinite Toeplitz matrix can be considered to be a subset of $\{1, 2, \ldots, n\}$. A pattern P is said to be *(positive semidefinite) completable* if every partial positive Toeplitz matrix with pattern P admits a positive semidefinite completion. We will use the following result.

Theorem 1.3.25 *A pattern $P \subseteq \{1, 2, \ldots, n\}$ is completable if and only if it is an arithmetic progression.*

Proof. The proof may be found in [336]. □

Although the statement above and Theorem 1.3.8 seem to be quite close, they are rather different. For instance, if we let $\Lambda = P = \{1, 2, 4\}$, then Theorem 1.3.8 yields that there exist c_k, $k \in \Lambda - \Lambda = \{-3, \ldots, 3\}$, such that

$$\begin{pmatrix} c_0 & c_{-1} & c_{-3} \\ c_1 & c_0 & c_{-2} \\ c_3 & c_2 & c_0 \end{pmatrix} \geq 0$$

and

$$\begin{pmatrix} c_0 & c_{-1} & c_{-2} & c_{-3} \\ c_1 & c_0 & c_{-1} & c_{-2} \\ c_2 & c_1 & c_0 & c_{-1} \\ c_3 & c_2 & c_1 & c_0 \end{pmatrix} \ngeq 0.$$

On the other hand, Theorem 1.3.25 yields in this case that there exist $c_k = \overline{c_k}$, $k = 0, 1, 2, 4$, so that the partial matrix

$$\begin{pmatrix} c_0 & c_{-1} & c_{-2} & ? & c_{-4} \\ c_1 & c_0 & c_{-1} & c_{-2} & ? \\ c_2 & c_1 & c_0 & c_{-1} & c_{-2} \\ ? & c_2 & c_1 & c_0 & c_{-1} \\ c_4 & ? & c_2 & c_1 & c_0 \end{pmatrix}$$

is partially positive semidefinite but no positive semidefinite Toeplitz completion exists.

Next, we will introduce a canonical form for a finite set $\Lambda \subset \mathbb{Z}^2$. We will be interested in those satisfying $G(\Lambda - \Lambda) = \mathbb{Z}^2$. Exercise 1.6.35 gives an easy method to determine whether a finite set $\Lambda \subset \mathbb{Z}^2$ satisfies $G(\Lambda - \Lambda) = \mathbb{Z}^2$ or not. By Exercise 1.6.30, a finite subset $\Lambda \subset \mathbb{Z}^2$ has the extension property if and only if every translation of Λ by a vector in $a \in \mathbb{Z}^2$ and every set isomorphic to Λ has the same property.

Let $d_1 = \max\{p_1 - p_2 : (p_1, q), (p_2, q) \in \Lambda\}$, $d_2 = \max\{q_1 - q_2 : (p, q_1), (p, q_2) \in \Lambda\}$, and $d = \max\{d_1, d_2\}$. By using a translation and maybe also interchanging the order of coordinates in $\mathbb{Z}^2$, we may assume that $(0, 0), (d, 0) \in \Lambda$, and that for each $k < 0$, $\max\{p_1 - p_2 : (p_1, k), (p_2, k) \in \Lambda\} < d$. If Λ has the above properties, then we say it is in *canonical form*. This notion plays a crucial role in our considerations. For Λ in canonical form, let $m = -\min\{l : (k, l) \in \Lambda\}$, and $M = \max\{l : (k, l) \in \Lambda\}$. Without loss of generality, we assume that $M \geq m$ for every Λ in canonical form.

If Λ is in canonical form and $M = m = 0$, then Λ is isomorphic to a subset of $\mathbb{Z}$ for which we may apply Theorem 1.3.8. Therefore we may assume that $M \geq 1$.

Lemma 1.3.26 *Suppose that Λ is in canonical form and has the extension property. Let $\Sigma = \{k : (k,0) \in \Lambda - \Lambda\}$ and $\Lambda(l) = \{(k,l) \in \Lambda\}$, $l \in \mathbb{Z}$. Then Σ forms an arithmetic progression. Moreover, there exists an l such that $\Sigma = \{k : (k,0) \in \Lambda(l) - \Lambda(l)\}$.*

Proof. Let $N = \text{card}\Sigma$. Define the sequence $\{c_{k,l}\}_{(k,l)\in\Lambda-\Lambda}$ by $c_{k,l} = c_k$ whenever $l = 0$ and $c_{k,l} = 0$ whenever $l \neq 0$. The sequence c_k will be specified later in the proof. The matrix of the sequence $\{c_{k,l}\}_{(k,l)\in\Lambda-\Lambda}$ can be written in a block diagonal form with $m + M + 1$ diagonal blocks (one for each row in Λ). Let Θ_j, $j = -m, \ldots, M$ denote these diagonal blocks. All the entries of the matrices Θ_j are terms of the sequence $\{c_k\}$. The sequence $\{c_{k,l}\}_{(k,l)\in\Lambda-\Lambda}$ is positive semidefinite if and only if all matrices Θ_j are positive semidefinite. The matrices Θ_j can be viewed also as fully specified principal submatrices of the partial Toeplitz matrix $(c_{s-t})_{s,t=0}^{d}$. If $\{k_j\}_{j=1}^{N} = \Sigma$ is not an arithmetic progression, then by Theorem 1.3.25 we can choose c_k such that all the matrices Θ_j are positive semidefinite but the sequence c_k does not have a positive semidefinite Toeplitz extension. This implies that the sequence $\{c_{k,l}\}$ constructed earlier in the proof is positive semidefinite and does not have a positive semidefinite extension. This proves that Σ must be an arithmetic progression.

Each Θ_l is an $n(l) \times n(l)$ matrix, where $n(l) = \text{card } \Lambda(l)$ and $c_{k,0}$ is an entry of Θ_l if and only if $(k,0) \in \Lambda(l) - \Lambda(l)$. Let $n = \max_l n(l)$. Then we can redefine the sequence $\{c_k\}$ used above with $c_0 = 1$ and $c_k = -1/(n-1)$ if $k \neq 0$. Then each matrix Θ_j is positive semidefinite by diagonal dominance. It can easily be verified that if $N > n$ the sequence $\{c_k\}$ is not positive semidefinite. Therefore we must have $n = N$. But then the theorem follows since $\Lambda(l) - \Lambda(l) \subseteq \Sigma$ for all l. $\square$

Lemma 1.3.27 *Suppose that Λ is in canonical form, $G(\Lambda - \Lambda) = \mathbb{Z}^2$, and Λ has the extension property. Then there exist p and q such that*

$$(p,q), (p+1,q), \ldots, (p+d,q) \in \Lambda.$$

In the remainder of the section $\{(p,q), (p+1,q), \ldots, (p+d,q)\}$ will be referred to as the *full row*.

Proof. The lemma is immediate for $d \leq 1$, so we assume that $d > 1$.

Let k be such that $(k,M) \in \Lambda$. Since the polynomial

$$1 + w^d + z^M w^k = 0 \tag{1.3.18}$$

has $2Md$ zeros on $\mathbb{T}^2$, then by Theorem 1.3.16, Λ does not have the extension property unless it contains more than $2Md$ points.

Assume that no row Λ contains $d+1$ elements. Then the maximum number of elements in Λ is $(M+m+1)d$. The inequality $(M+m+1)d \leq 2Md$ is true if $m \leq M-1$, therefore the lemma is certainly true unless $M = m$.

Let q be a row of Λ containing the largest number of elements among all rows. If $\{(l_0,q),(l_1,q),\dots,(l_m,q)\}$ is the set of elements on this row, then by Lemma 1.3.26, $l_0,l_1,\dots,l_m$ must be an arithmetic progression. Therefore, if Λ does not have a full row, each row of Λ contains at most $\frac{d}{2}+1$ elements. All rows $q<0$ contain then at most $\frac{d}{2}$ elements. Assuming $M=m$, we get that the maximal numbers of elements in Λ is $(M+1)(\frac{d}{2}+1)+M\frac{d}{2}$. Using again the polynomial in (1.3.18), we have that Λ does not have the extension property unless $(M+1)(\frac{d}{2}+1)+M\frac{d}{2}\le 2Md$, which holds if and only if

$$\frac{M+1}{M-\frac{1}{2}}\le d.$$

The function $f(M)=\frac{M+1}{M-\frac{1}{2}}$ is strictly decreasing. Since $f(2)=2$, the inequality holds for any $d>1$ if $M\ge 2$. It remains to consider the case when $M=1$. Since $f(1)=4$, we can restrict the consideration to the case when $d=2$ or $d=3$. If $d=3$, there will be at most 2 points in each row by Lemma 1.3.26 unless there is a row with four elements. There are at most 5 elements in Λ in this case. By counting the number of zeros on $\mathbb{T}^2$ of the polynomial (1.3.18), this possibility is ruled out. If $d=2$, the same reasoning holds if there are fewer than 5 elements in Λ. On the other hand, if there are 5 elements in Λ, three of them must be $(k,1)$, $(k+2,1)$, and $(w,-1)$. Since the polynomial $z^kw+z^{k+2}w+z^lw^{-l}=0$ has 8 zeros on $\mathbb{T}^2$, the proof is complete. □

Lemma 1.3.28 *Suppose $\{(0,0),(1,0)\}\subset\Lambda$ and that Λ is in canonical form. Then Λ does not have the extension property unless it contains elements of the form (p_j,j) for every j such that $-m\le j\le M$.*

Proof. Suppose there exists $-m\le q\le M$ such that there are no elements of the form (p,q) in Λ.

Since $G(\Lambda-\Lambda)=\mathbb{Z}^2$ it is clear that the set $Q=\{q:(p,q)\in\Lambda \text{ for some } p\}$ is not an arithmetic progression. Then it follows from Theorem 1.3.8 that there exists a data set $\{c_j, j\in Q-Q\}$, where $c_0=1$ without loss of generality, such that

$$A=\left(c_{q_1-q_2}\right)_{q_1,q_2\in Q}$$

is positive semidefinite, but its elements are not extendable to a positive semidefinite sequence on the set $[-m,M]$.

Let $J_{k\times k}$ be the $k\times k$ matrix with all entries entries equal to one, where

$$k=\max_{j,j_1,j_2}\{\,|p_{j_1}-p_{j_2}|:(p_{j_1},j),(p_{j_2},j)\in\Lambda\,\}.$$

The *Kronecker product* $B=A\otimes J_{k\times k}$ is a positive semidefinite matrix. Moreover, there exists a principal submatrix of B which gives a positive semidefinite sequence $(c_k)_{k\in\Lambda-\Lambda}$.

Let $c_{1,0}=c_{0,0}=1$. Since

$$\begin{pmatrix} c_{0,0} & c_{1,0} & c_{p_j,j}\\ \overline{c}_{1,0} & c_{0,0} & c_{p_j-1,j}\\ \overline{c}_{p_j,j} & \overline{c}_{p_j-1,j} & c_{0,0}\end{pmatrix}$$

must be positive semidefinite for any positive semidefinite sequence on $\mathbb{Z}^2$, it is clear that for all extensions of $(c_{j,l})_{(j,l)\in\Lambda-\Lambda}$ the $c_{j,l}$ must be constant along each row. Since the elements of A are not extendable to a positive semidefinite sequence on $\mathbb{Z}$, it follows that we have constructed a positive sequence on $\Lambda-\Lambda$ which is not extendable to a positive sequence on $\mathbb{Z}^2$. $\square$

Lemma 1.3.29 *Suppose Λ is a canonical form with $d=1$. Then Λ is homomorphic to $R(1,1)$, to $R(1,1)\setminus(1,1)$, to a set Λ' in canonical form with $d>1$, or to a set $\Lambda=\{(0,0),(1,0),(0,1),(p,q)\}$ which does not have the extension property.*

Proof. By Lemma 1.3.28 there exists an element $(r,1)\in\Lambda$; it is possible that $(r+1,1)\in\Lambda$. The isomorphism Φ defined by the matrix

$$\begin{pmatrix} 1 & -r \\ 0 & 1 \end{pmatrix} \tag{1.3.19}$$

maps Λ onto a set Δ which contains the elements $(0,0),(1,0),(0,1)$, and possibly also $(1,1)$. If Λ does not contain any additional element the proof is complete. Therefore, we may assume that Λ contains another element which is mapped to $(p,q)\in\Delta$ by Φ. If $(1,1)\in\Delta$ and $\mathrm{card}\Delta>4$, then Δ, or a translation of it, is isomorphic to a set containing an element $(d,0)$, $d\geq 2$. The latter follows by Exercise 1.6.36 since $\gcd((p,q)-(i,j))\geq 2$ for at least one choice of (i,j) among the elements $(0,0),(1,0),(0,1),(1,1)$. This proves the lemma if $(1,1)\in\Delta$.

If $S=\{(0,0),(1,0),(0,1),(p_1,q_1),(p_2,q_2)\}\subseteq\Delta$ it is clear that there exist $s_1,s_2\in S$ such that $\gcd\{p,q\}\geq 2$ where $(p,q)=s_1-s_2$. Then by Exercise 1.6.36, Δ is isomorphic to a set Λ' with $d>1$.

It remains to discuss the case when $\Delta=\{(0,0),(1,0),(0,1),(p,q)\}$. If $(p,q)\in\{(1,1),(1,-1),(-1,1),(-1,-1)\}$, then Δ is isomorphic to $R(1,1)$. If $(p,q)\notin\{(1,1),(1,-1),(-1,1),(-1,-1)\}$, then Δ does not have the extension property. This follows from the fact that $c_{0,0}=c_{1,0}=c_{0,1}=c_{-1,1}=1$, and $c_{p,q}=c_{p-1,q}=c_{p,q-1}=0$ is a positive semidefinite sequence without a positive semidefinite extension. $\square$

Remark 1.3.30 If the set Λ is in canonical form, there is a maximal j such that (p_0,j) and (p_d,j) are elements in Λ with $p_d-p_0=d$. We may assume that $j\leq M-m$. Otherwise translate (p_0,j) into $(0,0)$, and change the sign of the first coordinate of the indices. Then we obtain a new set in canonical form. For this new set we have $M_1=m+j$, and the maximal j has the same value as in the original set. We want to maximize M, and since we are free to work with any of the two sets introduced, we may assume $M\geq M_1$, i.e. $j\leq M-m$. Therefore, for Λ in canonical form we may assume that $\mathrm{card}\Lambda\leq(j+1)(d+1)+(M-j+m)d$.

The above remark leads to the following lemma.

Lemma 1.3.31 *Suppose Λ is in canonical form with $d>1$. Then Λ does not have the extension property if $M-m>1$ and $\max\{M-m,d\}\geq 3$.*

Proof. The polynomial (1.3.18) has $2Md$ zeros on $\mathbb{T}^2$. By Theorem 1.3.16, Λ does not have the extension property unless $\mathrm{card}\Lambda \leq 2Md$. Λ has at most $(j+1)(d+1)+(M-j+m)d$ elements. The latter number is maximized by $j = M-m$, and we get the inequality $(M-m+1)(d+1)+(M-(M-m)+m)d \leq 2Md$, which can be rewritten as

$$\frac{M-m+1}{M-m-1} \leq d.$$

Let $k = M-m$, and consider the function $f(k) = (k+1)/(k-1)$. Since this function is strictly decreasing and $f(2) = 3$ and $f(3) = 2$, the result follows. □

Lemma 1.3.32 *Suppose Λ is a set in canonical form with $m = 0$, $M = 1$, and $d > 1$. Then Λ has the extension property if and only if it is isomorphic to one of the sets $R(1,d)$ or $R(1,d) \setminus \{(1,d)\}$.*

Proof. By Lemma 1.3.27 at least one of the rows must be full. Let us assume, without loss of generality, that it is the zeroth row. Furthermore, using the isomorphism (1.3.19) we may assume that $(0,1) \in \Lambda$. Using the polynomial (1.3.18) with $M = 1$ and $p_M = 0$ and Theorem 1.3.16 shows that Λ does not have the extension property unless its cardinality is at least $2d+1$. Then $\Lambda = R(1,d) \setminus \{(1,d)\}$, if $(d,1) \notin \Lambda$. If $(d,1) \in \Lambda$, the two polynomials $1 - z^d w = 0$ and $z^d - w = 0$ have $2d$ common zeros on $\mathbb{T}^2$. Then Theorem 1.3.16 and the definition of the canonical form imply that Λ does not have the the extension property unless $\Lambda = R(1,d)$. The proof of the lemma is concluded by the fact that $R(1,d)$ and $R(1,d) \setminus (1,d)$ have the extension property. □

Lemma 1.3.33 *Let Λ be in canonical form with $M + m \geq 2$ and $d > 1$. If there exist p and q, $p \neq q$, such that $(p,M),(q,M) \in \Lambda$ or $(p,-m),(q,-m) \in \Lambda$, then Λ does not have the extension property.*

Proof. If $m = 0$, then Lemma 1.3.31 implies that Λ does not have the extension property except maybe for $M = d = 2$. If $M = d = 2$ the polynomial (1.3.18) has 8 zeros on $\mathbb{T}^2$. This implies that if Λ has the extension property, then it contains at least 9 elements. If Λ contains 9 or more elements, there exists p_2 such that $\{(0,0),(2,0),(p_2,2),(p_2+2,2)\} \subset \Lambda$. The polynomials $1 - z^{p_2+2}w^2 = 0$ and $z^2 - w^2 z^{p_2} = 0$ have 8 common zeros on $\mathbb{T}^2$, and then Theorem 1.3.16 implies that Λ does not have the extension property. Therefore we may assume $m \geq 1$ for the rest of the proof.

We claim that if there are at most two elements in the Mth and $-m$th row of Λ, and for either $i = -m$ or M the elements in the ith row are (p_i, i) and $(p_i + k, i)$ with $k > \lfloor \frac{d}{2} \rfloor$, then Λ does not have the extension property. Using these two elements together with an element from the other extreme row we can construct a polynomial of a similar form as (1.3.18) with $2k(M+m)$ zeros on $\mathbb{T}^2$. Let j be as in 1.3.30. Then there are at most $4 + (j+1)(d+1) + (M-j+m-2)d$ elements in Λ. By Theorem 1.3.16, Λ

does not have the extension property if

$$4 + (j+1)(d+1) + (M - j + m - 2)d \leq 2k(M+m).$$

The left-hand side of the above inequality is maximized by $j = M - m$ which implies

$$(M + m - 1)d + M - m + 5 \leq 2k(M+m). \tag{1.3.20}$$

If d is even then $k \geq \frac{d}{2} + 1$. Using this lower bound for k, inequality (1.3.20) simplifies to

$$5 \leq d + M + 3m,$$

which holds since $d \geq 2$ and $m \geq 1$. If d is odd, $k \geq \frac{d}{2} + \frac{1}{2}$. Using this lower bound for k, inequality (1.3.20) simplifies to

$$5 \leq d + 2m,$$

which holds since $d \geq 3$ and $m \geq 1$. This proves the claim.

To complete the proof of the lemma we will produce a nonextendable positive semidefinite data set in the case when Λ contains a pair of elements of the form $(p_i, i), (p_i + k, i)$ with $i = -m$ or $i = M$ and $0 < k \leq \lfloor \frac{d}{2} \rfloor$. If there are more than two elements in row $-m$ or M, such a pair must exist. If there are at most two elements in these rows we may assume that such a pair exists by the previous claim. To produce a nonextendable positive data set we define $c_{0,0} = 1$, and $c_{p,q} = 0$ for any $(p,q) \neq (0,0)$ such that $q \neq M - (-m)$. Let us represent the matrix

$$\left(c_{p_1 - p_2, q_1 - q_2}\right)_{(p_1,q_1),(p_2,q_2) \in \Lambda}$$

in the block form

$$\begin{pmatrix} T_{M,M} & T_{M,M-1} & \cdots & \cdots & T_{M,-m} \\ & T_{M-1,M-1} & \ddots & & \vdots \\ & & \ddots & \ddots & \vdots \\ & & & T_{-m+1,-m+1} & T_{-m+1,-m} \\ & & & & T_{-m,-m} \end{pmatrix}, \tag{1.3.21}$$

where the block $T_{i,n}$ contains all the elements of the form $c_{p-q,i-n}$ for a fixed pair of second coordinates i and n. By our choice of values for $c_{p,q}$, all block diagonal elements are identity matrices, which sizes depend on the number of elements $(p,q) \in \Lambda$ for a fixed q. By Lemma 1.3.27 at least one of these block diagonal matrices is an $(d+1) \times (d+1)$ identity matrix. The off-diagonal blocks are, except for the $T_{M,-m}$ block, zero matrices of appropriate dimension. The dimension of both square block matrices $T_{M,M}$ and $T_{-m,-m}$ is less than $d+1$. The lemma now follows by choosing entries in $T_{M,-m}$ such that the matrix (1.3.21) is positive semidefinite, but when extending the matrices $T_{M,M}$, $T_{-m,-m}$, and $T_{M,-m}$ to $(d+1) \times (d+1)$ matrices we get a 2×2 block matrix which is not positive semidefinite. The matrices $T_{M,M}$ and $T_{-m,-m}$ are both submatrices of $(c_{i-n,0})_{i,n \in \{0,1,\ldots,d+1\}}$,

which equals the $(d+1)\times(d+1)$ identity matrix. The matrix $T_{M,-m}$ will be a submatrix of a Toeplitz matrix of the form

$$A = \begin{pmatrix} t_0 & t_1 & \dots & \dots & t_{\lfloor \frac{d}{2} \rfloor} \\ t_{-1} & t_0 & t_1 & \dots & t_{\lfloor \frac{d}{2} \rfloor -1} \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ t_{-\lfloor \frac{d}{2} \rfloor +1} & \dots & t_{-1} & t_0 & t_1 \\ t_{-\lfloor \frac{d}{2} \rfloor} & \dots & \dots & t_{-1} & t_0 \end{pmatrix}.$$

If Λ has the extension property, it must be possible to embed $T_{M,-m}$ into the bigger Toeplitz contraction of the form

$$T = \begin{pmatrix} t_0 & t_1 & \dots & \dots & t_d \\ t_{-1} & t_0 & t_1 & \dots & t_{d-1} \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ t_{-d+1} & \dots & t_{-1} & t_0 & t_1 \\ t_{-d} & \dots & \dots & t_{-1} & t_0 \end{pmatrix}.$$

The block matrix (1.3.21) is positive semidefinite if and only if $T_{M,-m}$ is a contraction. If $T_{M,-m}$ has only one column or one row, define $c_{p_i-n_1,i-n_2} = c_{p_{i+k}-n_1,i-n_2} = \frac{1}{\sqrt{2}}$ for $i = M$ and $n_2 = -m$, and all other entries equal to zero. If $i = -m$, make a similar construction. Then the information we are given about T prevents it from being a contraction. If $T_{M,-m}$ has more than one row and column, let $t_i = t_n = 1$, where n is the index of the lower left corner of $T_{M,-m}$ and i is the smallest index such that t_i is not an element of the first column or last row of $T_{M,-m}$. Let the other t_l of $T_{M,-m}$ be zero. Since $|i-n| \le d$ it is clear that this choice prevents T from being a contraction. □

Using the definition of canonical form, the following is an immediate consequence of Lemma 1.3.33.

Corollary 1.3.34 *Let Λ be such that $G(\Lambda-\Lambda) = \mathbb{Z}^2$, and suppose that Λ is in canonical form with $m = 0$, $M \ge 2$, and $d \ge 2$. Then Λ does not have the extension property.*

Lemma 1.3.35 *There is no set Λ in canonical form with $d = 2$ and $M = m = 1$ which has the extension property.*

Proof. By Lemma 1.3.27 and Lemma 1.3.33 it is sufficient to consider the case when Λ is of the form $\Lambda = \{(p_{-1},-1),(0,0),(1,0),(2,0),(p_1,1)\}$. Let $\{c_k\}_{k\in\Lambda-\Lambda}$ be a positive semidefinite sequence. The matrix of the sequence is of the form

$$\begin{pmatrix} c_{0,0} & c_{-p_1,-1} & c_{1-p_1,-1} & c_{2-p_1,-1} & c_{p_{-1}-p_1,-2} \\ & c_{0,0} & c_{1,0} & c_{2,0} & c_{p_{-1},-1} \\ & & c_{0,0} & c_{1,0} & c_{p_{-1}-1,-1} \\ & & & c_{0,0} & c_{p_{-1}-2,-1} \\ & & & & c_{0,0} \end{pmatrix}.$$

Let $c_{0,0} = 1$, $c_{1,0} = c_{2,0} = 0$, $c_{1-p_1,-1} = c_{2-p_1,-1} = \frac{1}{\sqrt{2}}$, $c_{k,-1} = 0$ for $k \neq 1 - p_1, 2 - p_1$, and

$$c_{p_{-1}-p_1,-2} = c_{-p_1,-1}c_{p_{-1},-1} + c_{1-p_1}c_{p_{-1}-1,-1} + c_{2-p_1,-1}c_{p_{-1}-2,-1}.$$

Then the matrix above is positive semidefinite. Still this data set does not have a positive extension as the Hermitian matrix

$$\begin{pmatrix} c_{0,0} & c_{1,0} & c_{1-p_1,-1} & c_{2-p_1,-1} \\ & c_{0,0} & c_{-p_1,-1} & c_{1-p_1,-1} \\ & & c_{0,0} & c_{1,0} \\ & & & c_{0,0} \end{pmatrix}$$

is not positive semidefinite. □

Lemma 1.3.36 *Suppose Λ is in canonical form with $d > 1$. In addition we assume that there are only one p such that $(p, M) \in \Lambda$ and only one q such that $(q, -m) \in \Lambda$. Then Λ does not have the extension property if $d + M - m \geq 3$.*

Proof. The proof follows the same lines as the proof of Lemma 1.3.31, but now there are at most $(j+1)(d+1) + (M - j + m - 2)d + 2$ elements in Λ. This number is maximized if $j = M - m$ and must be less than $2Md$. As a consequence we have the inequality

$$\frac{M - m + 3}{M - m + 1} \leq d.$$

The function $f(k) = (k+3)/(k+1)$ is strictly decreasing, and the lemma follows since $f(0) = 3$, $f(1) = 2$ and $f(2) < 2$. □

Corollary 1.3.37 *Every Λ in canonical form with $d \geq 2$, $m \geq 1$, and $d + M - m \geq 3$ does not have the extension property.*

Proof. The result follows from the definition of the canonical form, Lemma 1.3.33, and Lemma 1.3.36. □

Lemma 1.3.38 *Suppose Λ is in canonical form with $m \geq 2$ and Λ has a unique element on each of the rows $-m$, $-m+1$, $M-1$, and M. Then Λ does not have the extension property if $d > 1$.*

Proof. The proof is similar to the proof of Lemma 1.3.36. We have to consider in this case the inequality

$$(j+1)(d+1) + (M - j + m - 4)d + 4 \leq 2Md.$$

The left-hand side of the inequality is maximized when $j = M - m$. This leads to the inequality

$$\frac{M - m + 5}{M - m + 3} \leq d.$$

The function $f(k) = (k+5)/(k+3)$ is strictly decreasing. The lemma follows since $f(0) = 5/3 < 2$. □

Lemma 1.3.39 *Let Λ be in canonical form with $M+m \geq 4$ and $d > 1$. If Λ does not have a unique element on each of the rows $-m$, $-m+1$, $M-1$, and M, then Λ does not have the extension property.*

Proof. By a similar ordering of the elements as in the proof of Lemma 1.3.33 the matrix

$$\left(c_{p_1-p_2,q_1-q_2}\right)_{(p_1,q_1),(p_2,q_2)\in\Lambda}$$

can be written in the block form

$$\begin{pmatrix} T_{M,M} & T_{M,M-1} & \dots & \dots & T_{M,-m} \\ & T_{M-1,M-1} & \ddots & & \vdots \\ & & \ddots & \ddots & \vdots \\ & & & T_{-m+1,-m+1} & T_{-m+1,-m} \\ & & & & T_{-m,-m} \end{pmatrix}, \tag{1.3.22}$$

where the block $T_{i,j}$ contains all the elements of the form $c_{p-q,i-j}$ for fixed second coordinates i and j.

Define $c_{0,0} = 1$ and all other entries of the matrix which are not in the $T_{M,-m+1}$ or $T_{M-1,-m}$ blocks to be zero. By our choice of values for $c_{i,j}$, all block diagonal elements are identity matrices, whose sizes depend on the number of elements $(p,q) \in \Lambda$ for a fixed q. By Lemma 1.3.27, at least one of these block diagonal matrices is a $(d+1)\times(d+1)$ identity matrix. By Lemma 1.3.33, Λ does not have the extension property unless it contains exactly one element of both the forms (p,k), for $k = -m, M$. We can therefore assume this. This implies that both block matrices $T_{M,M}$ and $T_{-m,-m}$ are 1×1 matrices, and the $T_{M,-m+1}$ and $T_{M-1,-m}$ blocks are a single row and a single column, respectively. Both blocks $T_{M,-m+1}$ and $T_{M-1,-m}$ are submatrices of Toeplitz matrices (not necessarily the same one) of the form

$$T = \begin{pmatrix} t_0 & t_1 & \dots & \dots & t_d \\ t_{-1} & t_0 & t_1 & \dots & t_{d-1} \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ t_{-d+1} & \dots & t_{-1} & t_0 & t_1 \\ t_{-d} & \dots & \dots & t_{-1} & t_0 \end{pmatrix}.$$

This means that some entries in a row or column of T are specified. A positive semidefinite extension of the matrix (1.3.22) exists if and only if there exists a contractive extension of T.

Now, if $t_j = t_i = 1/\sqrt{2}$ with $|j-i| \leq \frac{d}{2}$ and $t_k = 0$ for $k \neq i,j$ the matrix (1.3.22) is a positive semidefinite matrix, but T cannot be extended to a contraction. If there are more than three elements of the form $(i_{M-1}, M-1) \in \Lambda$ (or of the form $(i_{-m+1}, -m+1) \in \Lambda$) or two elements with distance less than $d/2$ in row $M-1$ or $-m+1$, this construction is possible.

It remains to prove the lemma in the case when there are at most two elements in both rows $M-1$ and $-m+1$, and that the distance between them is more than $d/2$. By this assumption there are at most

$$2 + 4 + (M-1)(d+1) + (m-2)d = 5 + M + (M+m-3)d$$

elements in Λ. Moreover, there exists a polynomial of similar type as (1.3.18) with $2k(M+m-1)$ zeros on $\mathbb{T}^2$, where k is the distance between the elements in one of the rows $M-1$ or $-m+1$. We may assume that $k \geq \frac{d}{2}+\frac{1}{2}$ if d is odd, and $k \geq \frac{d}{2}+1$ if d is even. The proof is concluded by Theorem 1.3.16 since

$$M+5+(M+m-3)d \leq 2\left(\frac{d}{2}+\frac{1}{2}\right)(M+m-1)$$

holds if d is odd, $d \geq 3$, and $M \geq m \geq 2$, and

$$M+5+(M+m-3)d \leq 2\left(\frac{d}{2}+1\right)(M+m-1)$$

holds if d is even and $M \geq m \geq 2$. □

Corollary 1.3.40 *Suppose Λ is a canonical form with $d > 1$ and $m \geq 2$. Then Λ does not have the extension property.*

Proof. The result follows from Lemma 1.3.38 and Lemma 1.3.39 . □

d	m	$M-m$	
1	≥ 0	≥ 0	Lemma 1.3.29
2	0	$0,1$	Theorem 1.3.8 and Lemma 1.3.32
2	0	≥ 2	Corollary 1.3.34
2	1	0	Lemma 1.3.35
2	1	≥ 1	Corollary 1.3.37
2	≥ 2	≥ 0	Corollary 1.3.40
3	0	0,1	Theorem 1.3.8 and Lemma 1.3.32
3	0	≥ 2	Corollary 1.3.34
3	≥ 1	≥ 0	Corollary 1.3.37
≥ 4	0	0,1	Theorem 1.3.8 and Lemma 1.3.32
≥ 4	0	≥ 2	Corollary 1.3.34
≥ 4	≥ 1	≥ 0	Corollary 1.3.37

Table 1.1 Cases considered in the proof of Theorem 1.3.18.

Proof of Theorem 1.3.18. The sets Λ written in canonical form can be divided into the cases in Table 1.1, depending on the size d, m, and $M-m$. The table indicates which of the results cover the particular case under consideration. □

1.4 DETERMINANT AND ENTROPY MAXIMIZATION

In this section we study the following functions, defined on the interior of the cones PSD_n and $\mathbb{T}\mathrm{Pol}_n^+$, respectively:

$$\log\det : \mathrm{int}(\mathrm{PSD}_n) \to \mathbb{R}, \tag{1.4.1}$$

$$\mathcal{E} : \mathrm{int}(\mathbb{T}\mathrm{Pol}_n^+) \to \mathbb{R},\ \mathcal{E}(p) = \frac{1}{2\pi}\int_0^{2\pi} \log p(e^{it})dt. \tag{1.4.2}$$

These functions will be useful in the study of these cones. We will show that both functions are strictly concave on their domain and we will obtain optimality conditions for their unique maximizers. These functions are examples of *barrier functions* as they tend to minus infinity when the argument approaches the boundary of the domain.

1.4.1 The log det function

Let PD_n denote the open cone of $n \times n$ positive definite matrices, or equivalently, PD_n is the interior of PSD_n. In other words,

$$\mathrm{PD}_n = \{A \in \mathbb{C}^{n\times n} \ : \ A > 0\}.$$

Let $F_0, \ldots, F_m$ be $n\times n$ Hermitian matrices, and put $F(x) = F_0 + \sum_{i=1}^m x_i F_i$, where $x = (x_1, \ldots, x_m)^T \in \mathbb{R}^m$. We assume that $F_1, \ldots, F_m$ are linearly independent. Let $\Delta = \{x \in \mathbb{R}^m \ : \ F(x) \in \mathrm{PD}_n\}$, which we will assume to be nonempty. By a translation of the vector $(x_1, \ldots, x_n)$, we may assume without loss of generality that $0 \in \Delta$. Note that Δ is convex, as $F(x) > 0$ and $F(y) > 0$ imply that $F(sx + (1-s)y) = sF(x) + (1-s)F(y) > 0$ for $0 \le s \le 1$. Define now the function

$$\phi(x) = \log\det F(x), \quad x \in \Delta.$$

We start by computing its gradient and Hessian.

Lemma 1.4.1 *We have that*

$$\frac{\partial\phi}{\partial x_i}(x) = \mathrm{tr}(F(x)^{-1}F_i), \quad \frac{\partial^2\phi}{\partial x_i \partial x_j}(x) = -\,\mathrm{tr}(F(x)^{-1}F_iF(x)^{-1}F_j),$$

and consequently,

$$\nabla\phi(x) = \begin{pmatrix}\mathrm{tr}(F(x)^{-1}F_1) & \cdots & \mathrm{tr}(F(x)^{-1}F_m)\end{pmatrix}, \tag{1.4.3}$$

$$\nabla^2\phi(x) = \begin{pmatrix} -\,\mathrm{tr}\,F(x)^{-1}F_1F(x)^{-1}F_1 & \cdots & -\,\mathrm{tr}\,F(x)^{-1}F_1F(x)^{-1}F_m \\ \vdots & & \vdots \\ -\,\mathrm{tr}\,F(x)^{-1}F_mF(x)^{-1}F_1 & \cdots & -\,\mathrm{tr}\,F(x)^{-1}F_mF(x)^{-1}F_m \end{pmatrix}. \tag{1.4.4}$$

Proof. Without loss of generality we take $x = 0$ (one can always do a translation so that 0 is the point of interest). Let us take $m = 1$ and let us

compute $\phi'(0)$. Note that

$$\phi(x) = \log\det(F_0(I + xF_0^{-1}F_1)) \tag{1.4.5}$$
$$= \log\det F_0 + \log\det(I + xF_0^{-1}F_1) \tag{1.4.6}$$
$$= \log\det F_0 + \log(1 + x\operatorname{tr}(F_0^{-1}F_1) + O(x^2)) \tag{1.4.7}$$
$$= \log\det F_0 + x\operatorname{tr}(F_0^{-1}F_1) + O(x^2) \quad (x \to 0). \tag{1.4.8}$$

But then $\phi'(0) = \lim_{x\to 0} \frac{\phi(x)-\phi(0)}{x} = \operatorname{tr}(F_0^{-1}F_1)$. The general expression for $\frac{\partial\phi}{\partial x_i}(x)$ follows now easily. The second derivatives are obtained similarly. □

Corollary 1.4.2 *The function* $\log\det$ *is strictly concave on* PD_n.

Proof. Note that the entries in the Hessian of ϕ can be rewritten as

$$-\operatorname{tr}(F(x)^{-1/2}F_iF(x)^{-1/2}F(x)^{-1/2}F_jF(x)^{-1/2}).$$

Thus for $y \in \mathbb{R}^m$ we have that

$$y^*\nabla^2\phi(x)y = -\sum_{i,j=1}^{m} y_i \operatorname{tr}(F(x)^{-1/2}F_iF(x)^{-1/2}F(x)^{-1/2}F_jF(x)^{-1/2})y_j \tag{1.4.9}$$
$$= -\operatorname{tr}(F(x)^{-1/2}\left(\sum_{i=1}^{m} y_iF_i\right)F(x)^{-1/2})^2 \tag{1.4.10}$$
$$= -\|F(x)^{-1/2}\left(\sum_{i=1}^{m} y_iF_i\right)F(x)^{-1/2}\|_2 \le 0, \tag{1.4.11}$$

and equality holds if and only if $\sum_{i=1}^m y_iF_i = 0$. Here $\|\cdot\|_2$ stands for the *Frobenius norm of a matrix*, i.e., $\|A\|_2 = \sqrt{\operatorname{tr}(A^*A)}$. As $F_1, \ldots, F_m$ are linearly independent, this happens only if $y_1 = \cdots = y_m = 0$. □

The following is the main result of this subsection.

Theorem 1.4.3 *Let* $\mathcal{W}$ *be a linear subspace of* $\mathcal{H}_n$*, with* 0 *the only positive semidefinite matrix contained in it. For every* $A > 0$ *and* $B \in \mathcal{H}_n$*, there is a unique* $F \in (A + \mathcal{W}) \cap \mathrm{PD}_n$ *such that* $F^{-1} - B \perp \mathcal{W}$*. Moreover,* F *is the unique maximizer of the function*

$$f(X) = \log\det X - \operatorname{tr}(BX)\ ,\ X \in (A + \mathcal{W}) \cap \mathrm{PD}_n. \tag{1.4.12}$$

If A *and* B *are real matrices then so is* F.

Using the Hahn-Banach separation theorem one can easily show that when 0 is the only positive semidefinite matrix in $\mathcal{W}$, for every Hermitian B we have that $(B + \mathcal{W}^\perp) \cap \mathrm{PSD}_n \neq \emptyset$. Indeed, if the set is empty then the separation theorem yields the existence of a Hermitian Φ and a real number α such that $\operatorname{tr}(P\Phi) > \alpha$ for all $P > 0$ and $\operatorname{tr}(X\Phi) \le \alpha$ for all $X \in B + \mathcal{W}^\perp$. Since the positive definite matrices form a cone, it is easy to see that we must have $\alpha \le 0$, and since $\mathcal{W}^\perp$ is a subspace it is easy to see that we must

have $\Phi \in \mathcal{W}$. But then it follows that $\Phi \geq 0$ and $\Phi \neq 0$. Thus the content of the theorem is not weakened when one restricts oneself to $B > 0$.

Proof. First note that $(A + \mathcal{W}) \cap \mathrm{PD}_n$ is convex. Next, since 0 is the only positive semidefinite matrix in $\mathcal{W}$, $(A+\mathcal{W}) \cap \mathrm{PD}_n$ is a bounded set. Indeed, as each nonzero $W \in \mathcal{W}$ has a negative eigenvalue, we have that for each $0 \neq W \in \mathcal{W}$ the set $(A+\mathbb{R}W)\cap\mathrm{PSD}_n$ is bounded. Let $g(W) = \max\{a : A+aW \in \mathrm{PSD}_n\}$. As $(A+\mathcal{W})\cap\mathrm{PSD}_n$ is convex, we obtain that g is a continuous function on the unit ball $\mathcal{B}$ in $\mathcal{W}$. Using now the finite dimensionality and thus the compactness of $\mathcal{B}$, we get that g attains a maximum on $\mathcal{B}$, M say. But then we obtain that we must have that $(A+\mathcal{W})\cap\mathrm{PD}_n \subset (A+\mathcal{W})\cap\mathrm{PSD}_n$ lies in $A + M\mathcal{B}$, which proves the boundedness. Next, by Corollary 1.4.2, log det is strictly concave on PD_n, and since $\mathrm{tr}(BX)$ is linear in X, $f(X)$ is strictly concave. Moreover, $f(X)$ tends to $-\infty$ when X approaches the boundary of $(A + \mathcal{W}) \cap \mathrm{PD}_n$ (as $\det X$ tends to 0 as X approaches the boundary). Thus f takes on a unique maximum on $(A + \mathcal{W}) \cap \mathrm{PD}_n$ for a matrix denoted by F, say.

Fix an arbitrary $W \in \mathcal{W}$. Consider the function $f_{F,W}(x) = \log\det(F + xW) - \mathrm{tr}(B(F+xW))$ defined in a neighborhood of 0 in $\mathbb{R}$. Then $f'_{F,W}(0) = 0$ (since f has its maximum at F). Similarly as in the proof of Lemma 1.4.1, we obtain that

$$f'_{F,W}(0) = \frac{(\det(I + xF^{-1}W))'}{\det(I + xF^{-1}W)}\Big|_{x=0} - (\mathrm{tr}(B(F + xW)))'|_{x=0}$$

$$= \mathrm{tr}(F^{-1}W) - \mathrm{tr}(BW) = \mathrm{tr}((F^{-1} - B)W) = 0.$$

Since W is an arbitrary element of $\mathcal{W}$ we have that $F^{-1} - B \perp \mathcal{W}$. Assume that $G \in (A + \mathcal{W}) \cap \mathrm{PD}_n$ and $G^{-1} - A \perp \mathcal{W}$. Then $f'_{G,W}(0) = 0$ for any $W \in \mathcal{W}$ and since f is strictly convex it follows that $G = F$. This proves the uniqueness of F.

In case A and B are real matrices, one can restrict attention to real matrices, and repeat the above argument. The resulting matrix F will also be real. □

Corollary 1.4.4 *Given is a positive definite matrix $A = (A_{ij})_{i,j=1}^n$, a Hermitian matrix $B = (B_{ij})_{i,j=1}^n$, and an $n\times n$ symmetric pattern P with associated graph G. Then there exists a unique positive definite matrix $F \in A+\mathcal{H}_G^\perp$ such that $(F^{-1})_{ij} = B_{ij}$ for every $(i,j) \notin P$. Moreover, F maximizes the function $f(X) = \mathrm{logdet}X - \mathrm{tr}(BX)$ over the set of all positive definite matrices $X \in A + \mathcal{H}_G^\perp$. In case A and B are real, F is also real.*

Proof. Consider in Theorem 1.4.3, $\mathcal{W} = \mathcal{H}_G^\perp$. Then evidently the only positive definite matrix in $\mathcal{W}$ is 0. By Theorem 1.4.3 there exists a unique $F \in (A + \mathcal{W}) \cap \mathrm{PD}_n$ such that $F^{-1} - B \perp \mathcal{W}$. For any $(k,j) \notin P$, consider the matrix $W_R^{(k,j)} \in \mathcal{W}$ having all its entries 0 except those in positions (k,j) and (j,k), which equal 1, and the matrix $W_I^{(k,j)}$ having i in position (k,j), $-i$ in position (j,k), and 0 elsewhere. The conditions

$\mathrm{tr}((F^{-1}-B)W_R^{(k,j)}) = \mathrm{tr}((F^{-1}-B)W_I^{(k,j)}) = 0$ imply that $(F^{-1})_{kj} = B_{kj}$ for any $(k,j) \in P$. □

Corollary 1.4.5 *Let P be a symmetric pattern with associated graph G and let A be a Hermitian matrix such that there exists a positive definite matrix $B \in A + \mathcal{H}_G^{\perp}$. Then among all such positive definite matrices B there exists a unique one, denoted $\widetilde{A}$, which maximizes the determinant. Also, $\widetilde{A}$ is the only positive definite matrix in $A + \mathcal{H}_G^{\perp}$ the inverse of which has 0 in all positions corresponding to entries $(i,j) \notin P$.*

Proof. Follows immediately from Corollary 1.4.4 when $B = 0$. □

Let us end this subsection with an example for Corollary 1.4.5.

Example 1.4.6 Find

$$\max_{x_1,x_2} \det \begin{pmatrix} 3 & 2 & x_1 & -1 \\ 2 & 3 & 1 & x_2 \\ \overline{x_1} & 1 & 3 & 1 \\ -1 & \overline{x_2} & 1 & 3 \end{pmatrix},$$

the maximum being taken over those x_1 and x_2 for which the matrix is positive definite. Notice that for $x_1 = x_2 = 0$ the matrix above is positive semidefinite by diagonal dominance (or by using the Geršgorin disk theorem). As one can easily check, for $x_1 = x_2 = 0$ the matrix is invertible (and thus positive definite). Consequently, the above problem will have a solution. By Corollary 1.4.4 the maximum is obtained at real x_1 and x_2, and thus we can suffice with optimizing over the reals. The following Matlab script solves the problem. The script is a *damped Newton's method* with α as damping factor. Notice that y and H below correspond to the gradient and the Hessian, respectively.

```
F0=[3 2 0 -1; 2 3 1 0; 0 1 3 1 ; -1 0 1 3];
F1=[0 0 1 0 ; 0 0 0 0; 1 0 0 0; 0 0 0 0];
F2=[0 0 0 0 ; 0 0 0 1; 0 0 0 0; 0 1 0 0];
x=[0 ; 0];
G=inv(F0);
y = [G(1,3); G(2,4)];
while norm(y) > 1e-7,
   H=[G(1,3)*G(3,1)+G(1,1)*G(3,3) G(1,4)*G(2,3)+G(2,1)*G(3,4) ; ...
   G(1,4)*G(2,3)+G(2,1)*G(3,4) G(2,4)*G(4,2)+G(2,2)*G(4,4)];
   v=H\y;
   delta = sqrt(v'*y);
   if delta < 1/4 alpha=1; else alpha=1/(1+delta); end;
   x=x+alpha*v;
   Fx=F0+x(1)*F1+x(2)*F2;
   G=inv(Fx);
   y = [G(1,3); G(2,4)];
```

end;

The solution we find is

$$\begin{pmatrix} 3 & 2 & 0.3820 & -1 \\ 2 & 3 & 1 & -0.3820 \\ 0.3820 & 1 & 3 & 1 \\ -1 & -0.3820 & 1 & 3 \end{pmatrix}$$

with a determinant equal to 29.2705.

1.4.2 The entropy function

Let

$$\mathbb{T}\mathrm{Pol}_n^{++} = \{p \in \mathbb{T}\mathrm{Pol}_n \ : \ p(z) > 0, z \in \mathbb{T}\}.$$

Equivalently, $\mathbb{T}\mathrm{Pol}_n^{++}$ is the interior of $\mathbb{T}\mathrm{Pol}_n^{+}$. Let $p^{(0)}, \ldots, p^{(m)}$ be elements of $\mathbb{T}\mathrm{Pol}_n$, and put $p_x = p^{(0)} + \sum_{i=1}^m x_i p^{(i)}$, where $x = (x_1, \ldots, x_m)^T \in \mathbb{R}^m$. We will assume that $p^{(1)}, \ldots, p^{(m)}$ are linearly independent. Let $\Delta = \{x \in \mathbb{R}^m \ : \ p_x \in \mathbb{T}\mathrm{Pol}_n^{++}\}$, which we assume to be nonempty. It is easy to see that Δ is convex. Define now the function

$$\psi(x) = \mathcal{E}(p_x) := \frac{1}{2\pi} \int_0^{2\pi} \log p_x(e^{it}) dt. \tag{1.4.13}$$

Lemma 1.4.7 *We have that*

$$\frac{\partial \psi}{\partial x_j}(x) = \frac{1}{2\pi} \int_0^{2\pi} \frac{p^{(j)}(e^{it})}{p_x(e^{it})} dt,$$

$$\frac{\partial^2 \psi}{\partial x_j \partial x_k}(x) = -\frac{1}{2\pi} \int_0^{2\pi} \frac{p^{(j)}(e^{it}) p^{(k)}(e^{it})}{p_x(e^{it})^2} dt.$$

Proof. Let us take $m = 1$ and let us compute $\psi'(0)$ (assuming that $0 \in \Delta$). Note that

$$\psi(x) = \frac{1}{2\pi} \int_0^{2\pi} \left[\log p^{(0)}(e^{it}) - \log \left(1 + x \frac{p^{(1)}(e^{it})}{p^{(0)}(e^{it})} \right) \right] dt.$$

Writing $\log(1+z) = z + O(z^2)$ $(z \to 0)$, one readily finds that

$$\psi'(0) = \lim_{x \to 0} \frac{\psi(x) - \psi(0)}{x} = \frac{1}{2\pi} \int_0^{2\pi} \frac{p^{(1)}(e^{it})}{p^{(0)}(e^{it})} dt.$$

The general expression for $\frac{\partial \psi}{\partial x_i}(x)$ follows now easily. The second derivatives are obtained similarly. □

Corollary 1.4.8 *The function $\mathcal{E}$ is strictly concave on $\mathbb{T}\mathrm{Pol}_n^{++}$.*

Proof. For $y \in \mathbb{R}^m$ we have that

$$y^T \nabla^2 \psi(x) y = -\frac{1}{2\pi} \int_0^{2\pi} \frac{(\sum_{j=1}^m y_j p^{(j)}(e^{it}))^2}{p_x(e^{it})^2} dt \leq 0,$$

and equality holds if and only if $\sum_{j=1}^m y_j p^{(j)} \equiv 0$. As $p^{(1)}, \ldots, p^{(m)}$ are linearly independent, this happens only if $y_1 = \cdots = y_m = 0$. □

The following is the main result of this subsection.

Theorem 1.4.9 *Let $\mathcal{W}$ be a linear subspace of $\mathbb{T}\mathrm{Pol}_n$, with $\mathcal{W} \cap \mathbb{T}\mathrm{Pol}_n^+ = \{0\}$. For every $p \in \mathbb{T}\mathrm{Pol}_n^{++}$, there is a unique $g \in (p + \mathcal{W}) \cap \mathbb{T}\mathrm{Pol}_n^{++}$ such that $\frac{1}{g} - h \perp \mathcal{W}$. Moreover, g is the unique maximizer of the function*

$$\alpha(q) = \mathcal{E}(q) + \langle q, h \rangle \ , \ q \in (p + \mathcal{W}) \cap \mathbb{T}\mathrm{Pol}_n^{++}. \tag{1.4.14}$$

If p and h have real coefficients then so does g.

The proof is analogous to that of Theorem 1.4.3. We leave this as an exercise for the reader (see Exercise 1.6.48).

Example 1.4.10 Let $p_x(z) = \frac{1}{2z^3} + \frac{2}{z^2} + \frac{\overline{x}}{z} + 6 + xz + 2z^2 + \frac{z^3}{2}$. Determine

$$\max_x \mathcal{E}(p_x).$$

As the given data are real it suffices to consider real x. In order to perform the maximization, we need a way to determine the Fourier coefficients of $\frac{1}{p_x}$. We do this by using Matlab's command `fft`. Indeed, by using the fast Fourier transform we obtain a vector containing the values of p_x on a grid of the unit circle. Taking this vector and entry wise computing the reciprocal yields a vector containing the values of $\frac{1}{p_x}$ on a grid on $\mathbb{T}$. Performing an inverse fast Fourier transform now yields an approximation of the Fourier coefficients of $\frac{1}{p_x}$. We can now execute the maximization:

```
p=[.5;2;0;6;0;2;.5];
p1=[0;0;1;0;1;0;0];
x=0;
hh=fft([zeros(61,1);p;zeros(60,1)]);
hh=ones(128,1)./hh;
hh2=ifft(hh);
y=hh2(64);
while abs(y)>1e-5,
   ll=fft([zeros(62,1);[1;0;2;0;1];zeros(61,1)]);
   ll=ll.*abs(hh).*abs(hh);
   ll=ifft(ll);
   H=ll(65);
   v=H\y;
   delta=sqrt(v'*y);
   if delta < 1/4 alpha=1; else alpha = 1/(1+delta); end;
   x=x+alpha*v;
   p=p+x*p1;
   hh=fft([zeros(61,1);p;zeros(60,1)]);
   hh=ones(128,1)./hh;
   hh2=ifft(hh);
   y=hh2(64);
end;
```

The solution we find is $x = 0.4546$.

1.5 SEMIDEFINITE PROGRAMMING

Many of the problems we have discussed in this chapter can be solved numerically very effectively using semidefinite programming (SDP). We will briefly outline some of the main ideas behind semidefinite programming and present some examples. There is a general theory on "conic programming" that goes well beyond the cone of positive semidefinite matrices, but to discuss this in that generality is beyond the scope of this book.

Let $F_0, \ldots, F_m$ be given Hermitian matrices, and let $F(x) = F_0 + \sum_{i=1}^m x_i F_i$. Let also $c \in \mathbb{R}^m$ be given, and write $x = (x_1, \ldots, x_m)^T \in \mathbb{R}^m$. We consider the problem

$$\underset{\text{subject to } F(x)\geq 0}{\text{minimize}} \quad c^T x \quad . \tag{1.5.1}$$

We will call this the *primal* problem. We will denote

$$p_* = \inf\{c^T x \ : F(x) \geq 0\}.$$

The primal problem has a so-called *dual problem*, which is the following:

$$\underset{\substack{\text{subject to } Z\geq 0 \\ \text{and } \operatorname{tr}(F_i Z)=c_i, i=1,\ldots,m}}{\text{maximize}} \quad -\operatorname{tr}(F_0 Z) \quad . \tag{1.5.2}$$

We will denote

$$d_* = \sup\{-\operatorname{tr}(F_0 Z) \ : Z \geq 0 \text{ and } \operatorname{tr}(F_i Z) = c_i, i = 1, \ldots, m\}.$$

We will call the set $\{x \ : \ F(x) \geq 0\}$ the *feasible set for the primal problem*, and the set $\{Z \geq 0 \ : \ \operatorname{tr}(F_i Z) = c_i, i = 1, \ldots, m\}$ the *feasible set for the dual problem*. We say that a (primal) feasible x is *strictly feasible* if $F(x) > 0$. Similarly, a (dual) feasible Z is *strictly feasible* if $Z > 0$. Observe that when x is primal feasible and Z is dual feasible, we have that

$$c^T x + \operatorname{tr}(F_0 Z) = \sum_{i=1}^m \operatorname{tr}(F_i Z) x_i + \operatorname{tr}(F_0 Z) = \operatorname{tr}(F(x) Z) \geq 0$$

as $Z, F(x) \geq 0$. Thus

$$-\operatorname{tr}(F_0 Z) \leq c^T x.$$

Consequently, when one can find feasible x and Z, one immediately sees that

$$-\operatorname{tr}(F_0 Z) \leq d_* \leq p_* \leq c^T x.$$

Consequently, when one can find feasible x and Z such that

$$\text{duality gap} := c^T x + \operatorname{tr}(F_0 Z)$$

is small, then one knows that one is close to the optimum. We state here without proof, that if a strict feasible x or a strict feasible Z exists, one has in fact that $d_* = p_*$, so in that case one will be able to make the duality gap arbitrary small.

The log det function, which we discussed in the previous section, has three very useful properties: (i) it is strictly concave, and thus any maximizer will be unique; (ii) it tends to minus infinity as the argument approaches the boundary, and thus it will be easy to stay within the set of positive definite matrices; and (iii) its gradient and Hessian are easily computable. These three properties do not change when we add a linear function to log det. This now leads to modified primal and dual problems. Let $\mu > 0$, and introduce

$$\underset{\text{subject to } F(x)>0}{\text{minimize}} \quad \mu \log\det F(x) + c^T x \tag{1.5.3}$$

$$\underset{\substack{\text{subject to } Z>0 \\ \text{and } \operatorname{tr}(F_i Z)=c_i, i=1,\ldots,m}}{\text{maximize}} \quad -\operatorname{tr}(F_0 Z) + \mu \log\det Z \,. \tag{1.5.4}$$

When strictly feasible x and Z exist, the above problems will have unique optimal solutions which may for instance be found via *Newton's algorithm.* If the duality gap for these solutions is small, we will have found an approximate solution to our original problem. Of course, finding the right parameter μ (or rather the right updates for $\mu \to 0$) and finding efficient algorithms to implement these ideas are still an art, especially as closer to the optimal value the matrix becomes close to a singular one. It is beyond the scope of this book to go into further detail on semidefinite programming. In the notes we will refer the reader to references on the subject.

We end this section with two examples. We will make use of Matlab's LMI (Linear Matrix Inequalities) Toolbox to solve these problems numerically. The LMI Toolbox uses semidefinite programming.

Example 1.5.1 Find

$$\max_{x_1,x_2} \lambda_{\min} \begin{pmatrix} 1 & 2 & x_1 & -1 \\ 2 & 1 & 1 & x_2 \\ \overline{x_1} & 1 & 1 & 1 \\ -1 & \overline{x_2} & 1 & 1 \end{pmatrix}. \tag{1.5.5}$$

We first observe that as the given data are all real, it suffices to consider real x_1 and x_2. Using Matlab's command `gevp` (generalized eigenvalue problem) command, we get the following.

```
F0=[1 2 0 -1; 2 1 1 0; 0 1 1 1; -1 0 1 1]
F1=[0 0 1 0 ; 0 0 0 0; 1 0 0 0 ; 0 0 0 0]
F2=[0 0 0 0 ; 0 0 0 1; 0 0 0 0; 0 1 0 0]
setlmis([])
X1 = lmivar(1,[1 1]);
X2 = lmivar(1,[1 1]);
lmiterm([1 1 1 0],-F0)
lmiterm([1 1 1 X1],-F1,1)
lmiterm([1 1 1 X2],-F2,1)
```

```
lmiterm([-1 1 1 0],eye(4))
lmis = getlmis
[alpha,popt]=gevp(lmis,1,[1e-10 0 0 0 0 ])
x1 = dec2mat(lmis,popt,X1)
x2 = dec2mat(lmis,popt,X2)
```

Running this script leads to $\alpha = 1.0000, x_1 = 1.0000, x_2 = -1.0000$. This leads to

$$\alpha I + \begin{pmatrix} 1 & 2 & x_1 & -1 \\ 2 & 1 & 1 & x_2 \\ \overline{x}_1 & 1 & 1 & 1 \\ -1 & \overline{x}_2 & 1 & 1 \end{pmatrix} = \begin{pmatrix} 2 & 2 & 1 & -1 \\ 2 & 2 & 1 & -1 \\ 1 & 1 & 2 & 1 \\ -1 & -1 & 1 & 2 \end{pmatrix} \geq 0,$$

with α optimal. So the answer to (1.5.5) is -1.

Example 1.5.2 Find the outer factorization of $p(z) = 2z^{-2} - z^{-1} + 6 - z + z^2$. In principle one can determine the roots of $p(z)$ and use the approach presented in the proof of Theorem 1.1.5 to compute the outer factor. Here we will present a different approach based on semidefinite programming that we will be able to use in the matrix values case as well (see Section 2.4). We are looking for $g(z) = g_0 + g_1 z + g_2 z^2$ so that

$$p(z) = |g(z)|^2, \ z \in \mathbb{T}. \tag{1.5.6}$$

If we introduce the matrix

$$F := (F_{ij})_{i,j=0}^2 = \begin{pmatrix} g_0 \\ g_1 \\ g_2 \end{pmatrix} \begin{pmatrix} \overline{g_0} & \overline{g_1} & \overline{g_2} \end{pmatrix},$$

we have that $F \geq 0$, $\operatorname{tr}(F) = |g_0|^2 + |g_1|^2 + |g_2|^2 = 6$, $F_{10} + F_{21} = g_1\overline{g_0} + g_2\overline{g_1} = -1$, and $F_{20} = g_2\overline{g_0} = 2$. Here we used (1.5.6). Thus we need to find a positive semidefinite matrix F with trace equal to 6, the sum of the entries in diagonal 1 equal to -1, and the (2,0) entry equal to 2. In addition, we must see to it that g does not have any roots in the open unit disk. One way to do this is to maximize the value $|g(0)| = |g_0|$ which is the product of the absolute values of the roots of g. As we have seen in the proof of Theorem 1.1.5, the roots of any factor g of $p = |g|^2$ are the results of choosing one out each pair $(\alpha, 1/\overline{\alpha})$ of roots of p. When we choose the roots that do not lie in the open unit disk, we obtain an outer g. This is exactly achieved by maximizing $|g_0|^2$. In conclusion, we want to determine

$$\max \ F_{00} \ \text{ subject to } \begin{cases} F = (F_{ij})_{i,j=0}^2 \geq 0, \\ \operatorname{tr} F = 6, \ F_{01} + F_{21} = -1, \ F_{20} = 2. \end{cases}$$

When we write a Matlab script to find the appropriate $F \geq 0$ we get the following:

```
F0=[6 -1 2; -1 0 0 ; 2 0 0]     % Notice that any
F1=[1 0 0; 0 -1 0 ; 0 0 0]      % affine combination
```

```
F2=[0 0 0; 0 1 0; 0 0 -1]      % F =F0 + x1*F1 + x2*F2 + x3*F3
F3=[0 1 0; 1 0 -1 ; 0 -1 0]    % satisfies the linear constraints.
setlmis([])
x1 = lmivar(1,[1 1]); % x1, x2, x3 are a real scalars
x2 = lmivar(1,[1 1]);
x3 = lmivar(1,[1 1]);
lmiterm([-1 1 1 0],F0)
lmiterm([-1 1 1 x1],F1,1) % these lmiterm commands
lmiterm([-1 1 1 x2],F2,1) % create the LMI
lmiterm([-1 1 1 x3],F3,1) % 0 ≤ F0 + x1*F1 + x2*F2 + x3*F3
lmis = getlmis
[copt,xopt] = mincx(lmis,[-1;0;0],[1e-6 0 0 0 0 ]) % minimize -x1
F=F0+xopt(1)*F1+xopt(2)*F2 + xopt(3)*F3
chol(F)
```

Executing this script leads to

$$F = \begin{pmatrix} 5.1173 & -0.7190 & 2.0000 \\ -0.7190 & 0.1010 & -0.2810 \\ 2.0000 & -0.2810 & 0.7817 \end{pmatrix},$$

which factors as vv^* with $v^* = \begin{pmatrix} 2.2621 & -0.3178 & 0.8841 \end{pmatrix}$. Thus $p(z) = |2.2621 - 0.3178z + 0.8841z^2|^2$, $z \in \mathbb{T}$. It is no coincidence that the optimal matrix $F(x)$ is of rank 1 (see Exercise 1.6.49).

1.6 EXERCISES

1 Prove Proposition 1.1.1.

2 For the following subsets of $\mathbb{R}^2$ (endowed with the usual inner product), check whether the following sets are (1) cones; (2) closed. In the case of a cone determine their dual and their extreme rays.

(a) $\{x, y) \in \mathbb{R}^2 \ : \ x, y \geq 0\}$.

(b) $\{x, y) \in \mathbb{R}^2 \ : \ xy > 0\}$.

(c) $\{x, y) \in \mathbb{R}^2 \ : \ 0 \leq x \leq y\}$.

(d) $\{x, y) \in \mathbb{R}^2 \ : \ x + y \geq 0\}$.

3 Prove Lemma 1.1.2. (For the proof of part (ii) one needs the following corollary of the Hahn-Banach separation theorem: given a closed convex set $A \subset \mathcal{H}$ and a point $h_0 \notin A$, then there exist a $y \in \mathcal{H}$ and an $\alpha \in \mathbb{R}$ such that $\langle h_0, y\rangle < \alpha$ and $\langle a, y\rangle \geq \alpha$ for all $a \in A$.)

4 Show that $p(z) = \sum_{k=-n}^{n} p_k z^k$ is real for all $z \in \mathbb{T}$ if and only if $p_k = \overline{p_{-k}}$, $k = 0, \ldots, n$.

5 Prove the uniqueness up to a constant factor of modulus 1 of the outer (co-outer) factor of a polynomial $p \in \mathbb{T}\mathrm{Pol}_n^+$.

6 Show that $\mathrm{PSD}_G = (\mathrm{PSD}_G)^*$ if and only if all the maximal cliques in the graph G are disjoint.

7 For the following matrices A and patterns P determine whether A belongs to each of $\mathrm{PPSD}_G, \mathrm{PSD}_G, (\mathrm{PSD}_G)^*$, and $(\mathrm{PPSD}_G)^*$. Here G is the graph associated with P.

(a) $P = \{(i, j) \ ; \ 1 \leq i, j \leq 3 \text{ and } |i - j| \leq 1\}$, and $A = \begin{pmatrix} 5 & 1 & 0 \\ 1 & 2 & \frac{1}{2} \\ 0 & \frac{1}{2} & 1 \end{pmatrix}$.

(b) $P = \{(i, j) \ ; \ 1 \leq i, j \leq 3 \text{ and } |i - j| \leq 1\}$, and $A = \begin{pmatrix} 1 & 1 & 0 \\ 1 & 1 & \frac{1}{2} \\ 0 & \frac{1}{2} & \frac{1}{4} \end{pmatrix}$.

(c) $P = (\{1, \ldots, 4\} \times \{1, \ldots, 4\}) \setminus \{(1, 3), (2, 4), (3, 1), (4, 2)\}$ and
$$A = \begin{pmatrix} 1 & 1 & 0 & 1 \\ 1 & 2 & 1 & 0 \\ 0 & 1 & 1 & -1 \\ 1 & 0 & -1 & 2 \end{pmatrix}.$$

(d) $P = \{(i,j)\ ;\ 1 \le i,j \le 5 \text{ and } |i-j| \ne 2\}$ and

$$A = \begin{pmatrix} 1 & 2 & 0 & 4 & 5 \\ 2 & 4 & 6 & 0 & 10 \\ 0 & 6 & 9 & 12 & 0 \\ 4 & 0 & 12 & 16 & 20 \\ 5 & 10 & 0 & 20 & 25 \end{pmatrix}.$$

(e) $P = \{(i,j)\ ;\ 1 \le i,j \le 3 \text{ and } |i-j| \le 1\}$, and $A = \begin{pmatrix} 1 & 1 & 0 \\ 1 & 2 & 1 \\ 0 & 1 & 1 \end{pmatrix}$.

(f) $P = \{(i,j)\ ;\ 1 \le i,j \le 3 \text{ and } |i-j| \le 1\}$, and $A = \begin{pmatrix} 1 & 1 & 0 \\ 1 & 1 & 1 \\ 0 & 1 & 1 \end{pmatrix}$.

(g) $P = (\{1,\dots,4\} \times \{1,\dots,4\}) \setminus \{(1,3),(2,4),(3,1),(4,2)\}$ and

$$A = \begin{pmatrix} 1 & 1 & 0 & 1 \\ 1 & 1 & 1 & 0 \\ 0 & 1 & 1 & 1 \\ 1 & 0 & 1 & 1 \end{pmatrix}.$$

8 Can a chordal graph have 0,1,2,3 simplicial vertices? If yes, provide an example of such a graph. If no, explain why not.

9 Let

$$C_n = \{A = (A_{ij})_{i,j=1}^n \in \mathrm{PSD}_n \ :\ A_{ii} = 1, i = 1,\dots,n\}$$

be the set of correlation matrices.

(a) Show that C_n is convex.

(b) Recall that $E \in C_n$ is an *extreme point* of C_n if $E \pm \Delta \in C_n$ implies that $\Delta = 0$. Show that if $E \in C_3$ is an extreme point, then E has rank 1.

(c) Show that the convex set of real correlation matrices $C_3 \cap \mathbb{R}^{3\times 3}$ has extreme points of rank 2. (In fact, they are exactly the rank 2 real correlation matrices whose off-diagonal entries have absolute value less than one.)

(d) Show that if $E \in C_n$ is an extreme point of rank k then $k^2 \le n$.

(e) Show that if $E \in C_n \cap \mathbb{R}^{n\times n}$ is an extreme point of rank k then $k(k+1) \le 2n$.

10 Let G be an undirected graph on n vertices. Let $\beta(G)$ denote the minimal number of edges to be added to G to obtain a chordal graph. For instance, if G is the four cycle then $\beta(G) = 1$. Prove the following.

(a) If A generates an extreme ray of $\mathrm{PSD}_G \cap \mathbb{R}^{n\times n}$, then $\mathrm{rank} A \leq \beta(G)+1$. (Hint: Let $\widehat{G}$ be a chordal graph obtained from G by adding $\beta(G)$ edges. Let $A \in \mathrm{PSD}_G \cap \mathbb{R}^{n\times n}$ be of rank $r > \beta(G)+1$. Write $A = \sum_{k=1}^r A_k$ with $\mathrm{rank} A_k = 1$ and $A_k \in \mathrm{PSD}_{\widehat{G}} \cap \mathbb{R}^{n\times n}$. Show now that there exist positive $c_1, \ldots, c_r$, not all equal, such that $\sum_{k=1}^r c_k A_k \in \mathrm{PSD}_G \cap \mathbb{R}^{n\times n}$. Conclude that A does not generate an extreme ray.)

(b) If A generates an extreme ray of PSD_G, then $\mathrm{rank} A \leq 2\beta(G)+1$.

(c) Show that the bound in (a) is not sharp. (Hint: Consider the ladder graph G on $2m$ vertices $\{1, \ldots, 2m\}$ where the edges are $(2i-1, 2i), i = 1, \ldots, m$, and $(2i-1, 2i+1), (2i, 2i+2), i = 1, \ldots, m-1$, and show that $\beta(G) = m-1$ while all extreme rays in $\mathrm{PSD}_G \cap \mathbb{R}^{n\times n}$ are generated by matrices of rank ≤ 2.)

11 Let G be the loop of n vertices. Show that $\mathrm{PSD}_G \cap \mathbb{R}^{n\times n}$ has extreme rays generated by elements of rank $n-2$, and that elements of higher rank do not generate an extreme ray.

12 Let G be the $(3,2)$ full bipartite graph. Show that PSD_G has extreme rays generated by rank 3 matrices, while all extreme rays in $\mathrm{PSD}_G \cap \mathbb{R}^{5\times 5}$ are generated by matrices of rank 2 or less.

13 This exercise concerns the dependance on Λ of $\mathbb{T}\mathrm{Pol}^2_\Lambda$.

(a) Find a sequence $\{p_n\}_{n=-3}^3$ for which is the matrix (1.3.1) is positive semidefinite but the matrix (1.3.2) is not.

(b) Find a trigonometric polynomial $q(z) = q_{-3}z^{-3} + \cdots + q_3 z^3 \geq 0$ such that $q(z) = |h(z)|^2$, where $h(z)$ is some polynomial $h(z) = h_0 + h_1 z + h_2 z^2 + h_3 z^3$, but $q(z)$ cannot be written as $|\widetilde{h}(z)|^2$ with $\widetilde{h}(z) = \widetilde{h}_0 + \widetilde{h}_1 z + \widetilde{h}_3 z^3$.

(c) Explain why the existence of the sequence $\{p_n\}_{n=-3}^3$ in (a) implies the existence of the trigonometric polynomial $q(z)$ in (b), and vice versa.

14 Let $T_2 = (c_{i-j})_{i,j=0}^2$ be positive semidefinite with $\dim \mathrm{Ker} T_2 = 1$. Show that $c_{-2} = \frac{1}{c_0}(c_{-1}^2 + (c_0^2 - |c_1|^2)e^{it})$ for some $t \in \mathbb{R}$.

15 Let $T_n = (c_{i-j})_{i,j=0}^n$ be positive definite, and define $c_{n+1} = \overline{c}_{-n-1}$ via

$$c_{n+1} := \begin{pmatrix} c_n & \ldots & c_1 \end{pmatrix} \begin{pmatrix} c_0 & \ldots & c_{-n-1} \\ \vdots & \ddots & \vdots \\ c_{n-1} & \ldots & c_0 \end{pmatrix}^{-1} \begin{pmatrix} c_1 \\ \vdots \\ c_n \end{pmatrix} + \frac{\det T_n}{\det T_{n-1}} e^{it_0}, \tag{1.6.1}$$

where $t_0 \in \mathbb{R}$ may be chosen arbitrarily. Show that $T_{n+1} = (c_{i-j})_{i,j=0}^{n+1}$ is positive semidefinite and singular.

16 Let μ be the measure given by $\mu(\theta) = \frac{1}{|e^{2i\theta}-2|^2}d\theta$. Find its moments c_k for $k = 0, 1, 2, 3, 4$, and show that $T_4 = (c_{i-j})_{i,j=0}^4$ is strictly positive definite.

17 Prove that the cone $\mathcal{D}_S$ in the proof of Theorem 1.3.5 is closed.

18 Consider the cone of positive maps $\Phi : \mathbb{C}^{n\times n} \to \mathbb{C}^{m\times m}$ (i.e., Φ is linear and $\Phi(\mathrm{PSD}_n) \subseteq \mathrm{PSD}_m$). Show that the maps $X \mapsto AXA^*$ and $X \mapsto AX^TA^*$ generate extreme rays of this cone.

19 Write the following positive semidefinite Toeplitz matrices as the sum of rank 1 positive semidefinite Toeplitz matrices. (Hint: any $n \times n$ rank 1 positive semidefinite Toeplitz matrix has the form

$$c\begin{pmatrix} 1 \\ \overline{\alpha} \\ \vdots \\ \overline{\alpha}^{n-1} \end{pmatrix}\begin{pmatrix} 1 & \alpha & \cdots & \alpha^{n-1} \end{pmatrix},$$

where $c > 0$ and $\alpha \in \mathbb{T}$.)

(a) $T = \begin{pmatrix} 2 & 1-i & 0 \\ 1+i & 2 & 1-i \\ 0 & 1+i & 2 \end{pmatrix}$.

(b) $T = \begin{pmatrix} 3 & -1+\sqrt{2}(1-i) & 1-2i \\ -1+\sqrt{2}(1+i) & 3 & -1+\sqrt{2}(1-i) \\ 1+2i & -1+\sqrt{2}(1+i) & 3 \end{pmatrix}$.

20 Let $c_j = \overline{c}_{-j}$ be complex numbers such that

$$T_n = (c_{j-k})_{j,k=0}^n$$

is positive semidefinite.

(a) Assume that T_n has a one-dimensional kernel, and let $(p_j)_{j=0}^n$ be a nonzero vector in this kernel. Show that $p(z) := \sum_{j=0}^n p_j z^j$ has all its roots on the unit circle $\mathbb{T}$.

(b) Show that

$$\det\begin{pmatrix} c_0 & \cdots & c_{-n+1} & 1 \\ \vdots & & \vdots & \vdots \\ c_n & \cdots & c_1 & z^n \end{pmatrix}$$

has the same roots as p.

(c) Show that the roots of p are all different (start with the cases $n = 1$ and $n = 2$).

(d) Denote the roots of p by $\alpha_1, \ldots, \alpha_n$. Let $g^{(j)}(z) = g_0^{(j)} + \cdots + g_n^{(j)} z^n$ be the unique polynomial of degree n such that

$$g^{(j)}(z) = \begin{cases} 1 & \text{when } z = \alpha_j, \\ 0 & \text{when } z = \alpha_k,\ k \neq j. \end{cases}$$

Put $\rho_j = v_j^* T_n v_j$, where $v_j = (g_k^{(j)})_{k=0}^n$. Show that $c_j = \sum_{k=1}^n \rho_k \alpha_k^j$ for $j = -n, \ldots, n$.

(e) Show how the above procedure gives a way to find a measure of finite support for the truncated trigonometric moment problem (see Theorem 1.3.6) in the case that T_n has a one-dimensional kernel.

(f) Adjust the above procedure to the case when T_n has a kernel of dimension $k > 1$. (Hint: apply the above procedure to T_{n-k+1}.)

(g) Adjust the above procedure to the case when T_n positive definite. (Hint: choose a $c_{n+1} = \overline{c}_{-n-1}$ so that T_{n+1} is positive semidefinite with a one-dimensional kernel; see Exercise 1.6.15.)

21 Given a finite positive Borel measure μ on $\mathbb{T}^d$ we consider integrals of the form

$$I(f) = \int_{\mathbb{T}^d} f d\mu.$$

A *cubature* (or *quadrature*) formula for $I(f)$ is an expression of the form

$$C(f) = \sum_{j=1}^n \rho_j f(\alpha_j),$$

with fixed nodes $\alpha_1, \ldots, \alpha_n \in \mathbb{T}^n$ and fixed weights $\rho_1, \ldots, \rho_n > 0$, such that $C(f)$ is a "good" approximation for $I(f)$ on a class of functions of interest. Show that if we express the positive semidefinite Toeplitz matrix

$$(\widehat{\mu}(\lambda - \nu))_{\lambda,\nu \in \Lambda}$$

as

$$\left(\sum_{j=1}^n \rho_j \alpha_j^{\lambda - \nu} \right)_{\lambda,\nu \in \Lambda}$$

(which can be done due to Theorem 1.3.5), then we have that

$$I(p) = C(p) \text{ for all } p \in \mathrm{Pol}_{\Lambda - \Lambda}.$$

22 Use the previous two exercises to find a cubature formula for the measure $\mu(z) = \frac{1}{2\pi i} \frac{dz}{z|2-z|^2}$ on $\mathbb{T}$ that is exact for trigonometric polynomials of degree 2.

23 Consider

$$A = \begin{pmatrix} 1 & \bar{t}_1 & \bar{t}_2 \\ t_1 & 1 & \bar{t}_1 \\ t_2 & t_1 & 1 \end{pmatrix}, \; B = \begin{pmatrix} 1 & \bar{t}_1 & \bar{t}_3 \\ t_1 & 1 & \bar{t}_2 \\ t_3 & t_2 & 1 \end{pmatrix}.$$

Assume that $|t_1| = 1$. Show that $A \geq 0$ if and only if $t_2 = t_1^2$ and that $B \geq 0$ if and only if $|t_2| \leq 1$ and $t_3 = t_1 t_2$.

24 For $\gamma \in [0, \pi]$, consider the cone $\mathcal{C}_\gamma$ of real valued trigonometric polynomials $q(z)$ on $\mathbb{T}$ of degree n so that $q(e^{i\theta}) \geq 0$ for $-\gamma \leq \theta \leq \gamma$.

(a) Show that $q \in \mathcal{C}_\gamma$ if and only if there exist polynomials p_1 and p_2 of degree at most n and $n-1$, respectively, so that

$$q(z) = |p_1(z)|^2 + \left(z + \frac{1}{z} - 2\cos\gamma\right)|p_2(z)|^2, \; z \in \mathbb{T}.$$

(Hint: Look at all the roots of $q(z)$ and conclude that the roots on the unit circle where $q(z)$ switches sign lie in the set $\{e^{i\theta} : \theta \in [\gamma, 2\pi - \gamma]\}$ and that there exist an even number of them. Notice that the other roots appear in pairs $\alpha, \frac{1}{\bar{\alpha}}$, so that $q(e^{i\theta}) = q_1(e^{i\theta}) \prod_{k=1}^{2r} \sin \frac{t - t_k}{2}$, where $t_1, \ldots, t_{2r} \in [\gamma, 2\pi - \gamma]$ and q_1 is nonnegative on $\mathbb{T}$.)

(b) Determine the extreme rays of the cone $\mathcal{C}_\gamma$.

(c) Show that the dual cone of $\mathcal{C}_\gamma$ is given by those real valued trigonometric polynomials $g(z) = \sum_{k=-n}^{n} g_k z^k$ so that the following Toeplitz matrices are positive semidefinite:

$$(g_{i-j})_{i,j=0}^{n} \geq 0 \quad \text{and} \quad (g_{i-j+1} + g_{i-j-1} - 2g_{ij}\cos\gamma)_{i,j=0}^{n-1} \geq 0.$$

25 Prove the first statement of the proof of Lemma 1.3.23.

26 Using positive semidefinite matrices, provide an alternative proof for the fact that

$$F(s,t) = s^2t^2(s^2 + t^2 - 1) + 1$$

is not a sum of squares. (Hint: suppose that $F(s,t) = \sum_{i=1}^{k} g_i(s,t)^2$. Show that $g_i(s,t) = a_i + b_i st + c_i s^2 t + d_i st^2$ for some real numbers a_i, b_i, c_i and d_i. Now rewrite $F(s,t)$ as

$$F(s,t) = \begin{pmatrix} 1 & st & s^2t & st^2 \end{pmatrix} \left(\sum_{i=1}^{k} \begin{pmatrix} a_i \\ b_i \\ c_i \\ d_i \end{pmatrix} \begin{pmatrix} a_i & b_i & c_i & d_i \end{pmatrix} \right) \begin{pmatrix} 1 \\ st \\ s^2t \\ st^2 \end{pmatrix}.$$

Thus we have written

$$F(s,t) = \begin{pmatrix} 1 & st & s^2t & st^2 \end{pmatrix} A \begin{pmatrix} 1 \\ st \\ s^2t \\ st^2 \end{pmatrix},$$

where A is a positive semidefinite real matrix. Now argue that this is impossible.)

27 For the following polynomials f determine polynomials g_i with real coefficients so that $f = \sum_{i=1}^{k} g_i^2$, or argue that such a representation is impossible. For those for which the representation is impossible, try to find a representation as above with g_i rational. (Hint: try $1 + x^2$, $1 + x^2 + y^2$, etc., as the numerator of g_i.)

(a) $f(x) = x^4 - 2x^3 + x^2 + 16$.

(b) $f(x, y) = 1 + x^2y^4 + x^4y^2 - 3x^2y^2$.

(c) $f(x, y) = 2x^4 + 2xy^3 - x^2y^2 + 5x^4$.

(d) $f(x, y) = x^2y^4 + x^4y^2 - 2x^3y^3 - 2xy^2 + 2x^2y + 1$.

(e) $f(x, y, z) = x^4 - (2yz + 1)x^2 + y^2z^2 + 2yz + 2$.

(f) $f(x, y) = \frac{1}{2}x^2y^4 - (2x^3 - \beta x^2)y^3 + [(\beta^2 + 2)x^4 + \beta x^3 + (\beta^2 + 9)x^2 + 2\beta x]y^2 - (8\beta x^2 - 4x)y + 4(\beta^2 + 1)$. (Hint: consider the cases $0 < |\beta| < 2$, $\beta = 0$, and $|\beta| \geq 2$, separately.)

28 Recall that $p(x, y, z)$ is a homogeneous polynomial of degree n if $p(ax, ay, az) = a^n p(x, y, z)$ for all x, y, z, and a.

(a) Show that if $p(x, y, z)$ is a homogeneous polynomial in three variables of degree 3, then

$$4p(0, 0, 1) + p(1, 1, 1) + p(-1, 1, 1) + p(1, -1, 1) + p(-1, -1, 1)$$

$$-2(p(1, 0, 1) + p(-1, 0, 1) + p(0, 1, 1) + p(0, -1, 1)) = 0.$$

(b) Show that if $f(x, y, z)$ is a sum of squares of homogeneous polynomials of degree 3, then

$$f(1, 1, 1) + f(-1, 1, 1) + f(1, -1, 1) + f(-1, -1, 1) + f(1, 0, 1)$$

$$+f(-1, 0, 1) + f(0, 1, 1) + f(0, -1, 1) - \frac{4}{5}f(0, 0, 1) \geq 0.$$

(c) Use (b) to show that the so-called Robinson polynomial

$$x^6 + y^6 + z^6 - (x^4(y^2 + z^2) + y^4(x^2 + z^2) + z^4(x^2 + y^2)) + 3x^2y^2z^2$$

is not a sum of squares of polynomials.

(d) Show that the Robinson polynomial only takes on nonnegative values on $\mathbb{R}^3$.

29 Theorem 1.3.8 may be interpreted in the following way. Given a nonempty finite $S \subseteq \mathbb{Z}$ of the form $S = \Lambda - \Lambda$, then $\mathcal{B}_S = \{p \ : \ T_{p,\Lambda} \geq 0\}$ if and only if $S = \{jb : j = -k, \ldots, k\}$ for some $k, b \in \mathbb{N}$. For general finite S satisfying $S = -S$ one may ask whether the equality

$$\mathcal{B}_S = \{p \ : \ T_{p,\Lambda} \geq 0 \text{ for all } \Lambda \text{ with } \Lambda - \Lambda \subseteq S\}$$

holds. Show that for the set $S = \{-4, -3, -1, 0, 1, 3, 4\}$ equality does not hold.

30 Let $h : \mathbb{Z}^d \to \mathbb{Z}^d$ be a group isomorphism, and let Λ be a finite subset of $\mathbb{Z}^d$. Show that Λ has the extension property if and only if $h(\Lambda)$ has the extension property.

31 Prove that when $a \in \mathbb{T}^d$ and $G(\Lambda - \Lambda) \neq \mathbb{Z}^d$, then $\Omega_\Lambda(\{a\})$ contains more than one element.

32 The entropy of a continuous function $f \in C(\mathbb{T}^d)$, $f > 0$ on $\mathbb{T}^d$ is defined as

$$E(f) = \frac{1}{(2\pi)^d} \int\limits_{[0,2\pi]^d} \log f(e^{i\langle k,t\rangle}) dt.$$

Let $n \geq 1$, and let $\Lambda = R(0,n)$, $R(1,n)$, or $R(1,n) \setminus \{(1,n)\}$, and $\{c_k\}_{k \in \Lambda - \Lambda}$ be a positive sequence of complex numbers. We call a function

$$f(e^{i\langle k,t\rangle}) = \sum_{k \in \mathbb{Z}^d} c_k(f) e^{i\langle k,t\rangle}, \quad f \in C(\mathbb{T}^d),$$

a *positive definite extension of the sequence* $\{c_k\}_{k \in \Lambda - \Lambda}$ if $f > 0$ on $\mathbb{T}^d$ and $c_k(f) = c_k$ for $k \in \Lambda - \Lambda$. Assume that $\{c_k\}_{k \in \Lambda - \Lambda}$ admits an extension $f \in C(\mathbb{T}^d)$, $f > 0$ on $\mathbb{T}^d$. Prove that among all positive continuous extensions of $\{c_k\}_{k \in \Lambda - \Lambda}$, there is a unique one denoted f_0 which maximizes the entropy. Moreover, f_0 is the unique positive extension of the form $\frac{1}{P}$, with $P \in \Pi_{\Lambda - \Lambda}$.

33 Let $\{c_{kl}\}_{(k,l) \in R(1,n) - R(1,n)}$ be a positive semidefinite sequence. Show that there exists a unique positive definite extension $\{d_{kl}\}$, $k \in \{-1, 0, 1\}$, $l \in \mathbb{Z}$ of $\{c_{kl}\}_{(k,l) \in R(1,n) - R(1,n)}$ which maximizes the entropy $E(\{d_{kl}\})$, and that this unique extension is given via the formula

$$F_0(e^{it}, e^{iw}) = \frac{f_0(e^{iw})[f_0^2(e^{iw}) - |f_1(e^{iw})|^2]}{|f_0(e^{iw}) - e^{it} f_1(e^{iw})|^2},$$

where $f_0(e^{iw}) = \sum_{j=-\infty}^{\infty} d_{0j} e^{ijw}$ and $f_1(e^{iw}) = \sum_{j=-\infty}^{\infty} d_{1j} e^{ijw}$.

34 Prove that every positive semidefinite sequence on $S = \{-1, 0, 1\} \times \mathbb{Z}$ with respect to $R_1 = \{0, 1\} \times \mathbb{N}$ admits a positive semidefinite extension to $\mathbb{Z}^2$.

35 Suppose $\Lambda = \{(p_i, q_i)_{i=1}^n\} \subset \mathbb{Z}^2$ such that

$$A = \begin{pmatrix} p_1 & p_2 & \ldots & p_n \\ q_1 & q_2 & \ldots & q_n \end{pmatrix}$$

has rank 2. Prove the following.

(a) There exist integer matrices P and Q with $|\det P| = |\det Q| = 1$ such that

$$PAQ = \begin{pmatrix} f_1 & 0 & 0 & \dots & 0 \\ 0 & f_2 & 0 & \dots & 0 \end{pmatrix} = F,$$

where $f_1 = \gcd\{p_1, \dots, p_n, q_1, \dots, q_n\}$ and

$$f_2 = \frac{1}{f_1} \gcd \left\{ \det \begin{pmatrix} p_i & p_j \\ q_i & q_j \end{pmatrix}, 1 \le i < j \le n \right\}.$$

(b) $G(\Lambda - \Lambda) = \mathbb{Z}^2$ if and only if $f_1 = f_2 = 1$.

36 Let $\Lambda = \{(p_i, q_i) : i = 1, \dots, n\} \subset \mathbb{Z}^2$ and define

$$d = \max\{\gcd(p_i, q_i) : i = 1, \dots, n\}. \tag{1.6.2}$$

Show that there exists an isomorphism $\Phi : \mathbb{Z}^2 \to \mathbb{Z}^2$ such that $(d, 0) \in \Phi(\Lambda)$. In addition, show that d is the largest number with this property.

(Hint: if the maximum is obtained at $i = 1$, there exist integers k, l such that $p_1 k - q_1 l = d$. Define Φ via the matrix

$$\begin{pmatrix} k & -l \\ \frac{q_1}{d} & -\frac{p_1}{d} \end{pmatrix}.)$$

37 Let $f(x) = \sum_{k=0}^{m} f_k x^k$ be a polynomial such that $f(x) \ge 0$, $x \in \mathbb{R}$, and $f_m \ne 0$.

(a) Show that $m = 2n$ for some $n \in \mathbb{N}_0$.

(b) Show that $f_k \in \mathbb{R}$, $k = 0, \dots, 2n$.

(c) Show that α is a root of f if and only if $\overline{\alpha}$ is a root of f, and show that if α is a real root, then it has even multiplicity.

(d) Show that $f(x) = |g(x)|^2$ for some polynomial g of degree n with possibly complex coefficients.

(e) Show that g above can be chosen so that all its roots have nonnegative (nonpositive) imaginary part, and that with that choice g is unique up to a constant of modulus 1.

(f) Show that $f = h_1^2 + h_2^2$ for some polynomials h_1 and h_2 of degree $\le n$ with real coefficients. (Hint: take g as above and let h_1 and h_2 be real valued on the real line so that $g = h_1 + ih_2$ on the real line.)

(g) Use ideas similar to those above to show that a polynomial $\psi(x)$ with $\psi(x) \ge 0$, $x \ge 0$, can be written as

$$\psi(x) = h_1(x)^2 + h_2(x)^2 + x(h_3(x)^2 + h_4(x)^2),$$

where h_1, h_2, h_3, h_4 are polynomials with real coefficients.

38 Define the $(n+1)\times(n+1)$ matrices Φ_n and Ψ_n such that

$$\Phi_n \begin{pmatrix} 1 \\ u \\ \vdots \\ u^n \end{pmatrix} = \begin{pmatrix} (1+iu)^0(1-iu)^n \\ (1+iu)^1(1-iu)^{n-1} \\ \vdots \\ (1+iu)^n(1-iu)^0 \end{pmatrix},$$

$$\Psi_n \begin{pmatrix} 1 \\ z \\ \vdots \\ z^n \end{pmatrix} = \begin{pmatrix} (i-iz)^0(1+z)^n \\ (i-iz)^1(1+z)^{n-1} \\ \vdots \\ (i-iz)^n(1+z)^0 \end{pmatrix}.$$

For instance, when $n=2$ we obtain the matrix

$$\Phi_2 = \begin{pmatrix} 1 & -2i & -1 \\ 1 & 0 & 1 \\ 1 & 2i & -1 \end{pmatrix}.$$

(i) Show that $\Psi_n\Phi_n = 2^n I_{n+1} = \Phi_n\Psi_n$.

(ii) Prove that if $H = (h_{i+j})_{i,j=0}^n$ is a Hankel matrix then $T = \Phi_n H \Phi_n^*$ is a Toeplitz matrix, and vice versa. (A *Hankel matrix* $H = (h_{ij})$ is a matrix that satisfies $h_{ij} = h_{i+1,j-1}$ for all applicable i and j.)

39 Let $\mathbb{R}\mathrm{Pol}_n = \{g \ : \ g(x) = \sum_{k=0}^n g_k x^k, g_k \in \mathbb{R}\}$ be the space of real valued degree $\leq n$ polynomials, and let $\mathbb{R}\mathrm{Pol}_n^+ = \{g \in \mathbb{R}\mathrm{Pol}_n \ : \ g(x) \geq 0 \text{ for all } x \in \mathbb{R}\}$ be the cone of nonnegative valued degree $\leq n$ polynomials. On $\mathbb{R}\mathrm{Pol}_n$ define the inner product

$$\langle g, h\rangle = \sum_{i=0}^n g_i h_i, \quad \text{for} \quad g(x) = \sum_{k=0}^n g_k x^k \quad \text{and} \quad h(x) = \sum_{k=0}^n h_k x^k.$$

For $h(x) = \sum_{i=0}^{2n} h_i x^i$ define the Hankel matrix

$$H_h = (h_{i+j})_{i,j=0}^n.$$

(a) Show that $\mathbb{R}\mathrm{Pol}_{2n+1}^+ = \mathbb{R}\mathrm{Pol}_{2n}^+$ (see also part (a) of Exercise 1.6.37).

(b) Show that for $g(x) = \sum_{k=0}^n g_k x^k$ and $h(x) = \sum_{k=0}^{2n} h_k x^k$ we have that

$$\langle g^2, h\rangle = \begin{pmatrix} \overline{g_0} & \cdots & \overline{g_n} \end{pmatrix} H_h \begin{pmatrix} g_0 \\ \vdots \\ g_n \end{pmatrix}.$$

(c) Show that the dual cone of $\mathbb{R}\mathrm{Pol}_{2n}^+$ is the cone

$$\mathcal{C} = \{h \in \mathbb{R}\mathrm{Pol}_{2n} \ : \ H_h \geq 0\}.$$

(d) Show that the extreme rays of $\mathbb{R}\mathrm{Pol}_{2n}^+$ are generated by the polynomials in $\mathbb{R}\mathrm{Pol}_{2n}^+$ that have only real roots.

(e) Show that the extreme rays of $\mathcal{C}$ are generated by those polynomials h for which H_h has rank 1. (This part may be easier to do after reading the proof of Theorem 2.7.9.)

40 Let $p(x) = p_0 + \cdots + p_{2n}x^{2n}$ with $p_j \in \mathbb{R}$.

(a) Show that there exist polynomials $g_j(x) = g_0^{(j)} + \cdots + g_n^{(j)}x^n$, $j = 1, \ldots, m$, with real coefficients such that $p = \sum_{j=1}^m g_j^2$, if and only if there exists a positive semidefinite matrix $F = (F_{jk})_{j,k=0}^n$ satisfying

$$\sum_{l=0}^{k} F_{k-l,l} = p_k, \quad k = 0, \ldots, 2n. \tag{1.6.3}$$

(b) Let

$$F_0 = \begin{pmatrix} p_0 & \frac{1}{2}p_1 & & \\ \frac{1}{2}p_1 & p_2 & \ddots & \\ & \ddots & \ddots & \frac{1}{2}p_{2n-1} \\ & & \frac{1}{2}p_{2n-1} & p_{2n} \end{pmatrix} \in \mathbb{C}^{(n+1)\times(n+1)}.$$

Show that $F = F^*$ satisfies (1.6.3) if and only if there is an $n \times n$ Hermitian matrix $Y = (Y_{jk})_{j,k=1}^n$ so that

$$F = F_0 + \begin{pmatrix} 0_{n\times 1} & iY \\ 0_{1\times 1} & 0_{n\times 1} \end{pmatrix} + \begin{pmatrix} 0_{1\times n} & 0_{1\times 1} \\ -iY & 0_{n\times 1} \end{pmatrix}.$$

Here $0_{j\times k}$ is the $j \times k$ zero matrix.

(c) Let

$$A = -i \begin{pmatrix} 0 & & & \\ 1 & 0 & & \\ & \ddots & \ddots & \\ & & 1 & 0 \end{pmatrix} \in \mathbb{C}^{n\times n}, B = -i \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix} \in \mathbb{C}^{n\times 1}.$$

Show that

$$\begin{pmatrix} 0_{1\times n} & 0_{1\times 1} \\ -iY & 0_{n\times 1} \end{pmatrix} = \begin{pmatrix} 0_{1\times 1} & 0_{1\times n} \\ YB & YA \end{pmatrix}.$$

(d) Assume that $p_0 > 0$ and write F_0 as

$$F_0 = \begin{pmatrix} G_{11} & G_{12} \\ G_{21} & G_{22} \end{pmatrix},$$

where $G_{11} = p_0$ and $G_{22} \in \mathbb{C}^{n\times n}$. Show that F in (b) is positive semidefinite if and only if

$$A^*Y + YA - (YB + G_{21})G_{11}^{-1}(B^*Y + G_{12}) + G_{22} \geq 0. \tag{1.6.4}$$

In addition, show that F is of rank 1 if and only equality holds in (1.6.4).

(e) The equation

$$P^*Y + YP - YQY + R = 0 \tag{1.6.5}$$

for known matrices P, $Q = Q^*$, $R = R^*$, and unknown $Y = Y^*$ is referred to as the *Algebraic Riccati Equation*. Show that this Riccati equation has a solution Y if and only if the range of the matrix $\left(\begin{smallmatrix} I_n \\ Y \end{smallmatrix}\right)$ is an invariant subspace of the so-called *Hamiltonian matrix*

$$H = \begin{pmatrix} -P^* & Q \\ R & P \end{pmatrix}.$$

(Hint: rewrite the Riccati equation as $\left(\begin{smallmatrix} -Y & I_n \end{smallmatrix}\right) H \left(\begin{smallmatrix} I_n \\ Y \end{smallmatrix}\right) = 0$.)

(f) For $p(x) = 1+4x+10x^2+12x^3+9x^4$, determine the appropriate Riccati equation and its Hamiltonian. Use the eigenvectors of the Hamiltonian to determine an invariant subspace of H of the appropriate form, and obtain a solution to the Riccati equation. Use this solution to determine a rank 1 positive semidefinite F satisfying (1.6.3), and obtain a factorization $p(x) = |g(x)|^2$. Notice that if one chooses the invariant subspace of the Hamiltonian by using eigenvectors whose eigenvalues lie in the right half-plane, the corresponding g has all its roots in the closed left half-plane

Remark: The Riccati equation is an important tool in control theory to determine factorizations of matrix-valued functions with desired properties (see also Section 2.7).

41 Consider

$$\mathcal{C} = \{A = (a_{ij})_{i,j=1}^3 \in \mathbb{R}^{3\times 3} \ : \ A = A^T \geq 0 \text{ and } a_{12} = a_{33}\}.$$

(a) Show that $\mathcal{C}$ is a closed cone.

(b) Show that for $x, y \in \mathbb{R}$, not both 0, the matrix

$$\begin{pmatrix} x^2 \\ y^2 \\ xy \end{pmatrix} \begin{pmatrix} x^2 & y^2 & xy \end{pmatrix} \tag{1.6.6}$$

generates an extreme ray of $\mathcal{C}$.

(c) Show that all extreme rays are generated by a matrix of the type (1.6.6). (Hint: the most involved part is to show that any rank 2 element in $\mathcal{C}$ does not generate an extreme ray. To prove this, first show that if

$$\begin{pmatrix} a_1 \\ a_2 \end{pmatrix}, \begin{pmatrix} b_1 \\ b_2 \end{pmatrix}, \begin{pmatrix} c_1 \\ c_2 \end{pmatrix} \in \mathbb{R}^2$$

are such that $a_1b_1 + a_2b_2 = c_1^2 + c_2^2$, then there exist α and x_1, x_2, y_1, y_2 in $\mathbb{R}$ so that

$$\begin{pmatrix} \cos\alpha & -\sin\alpha \\ \sin\alpha & \cos\alpha \end{pmatrix} \begin{pmatrix} a_1 & b_1 & c_1 \\ a_2 & b_2 & c_2 \end{pmatrix} = \begin{pmatrix} x_1^2 & y_1^2 & x_1y_1 \\ x_2^2 & y_2^2 & x_2y_2 \end{pmatrix}.$$

The correct α will satisfy $\tan(2\alpha) = \frac{a_1b_1 - a_2b_2 - c_1^2 + c_2^2}{a_1b_2 + a_2b_1 - 2c_1c_2}$.)

(d) Determine the dual cone of $\mathcal{C}$. What are its extreme rays?

42 Let $A = (A_{ij})_{i,j=1}^n$ be positive definite. Prove that there exists a unique positive definite matrix $F = (F_{ij})_{i,j=1}^n$ such that $F_{ij} = A_{ij}$, $i \neq j$, $\mathrm{tr}F = \mathrm{tr}A$ and the diagonal entries of the inverse of F are all the same. In case A is real, F is also real. (Hint: consider in Theorem 1.4.3 $B = 0$ and let $\mathcal{W}$ be the set of all real diagonal $n \times n$ matrices with their trace equal to 0.)

43 Consider the partial matrix

$$\mathcal{A} = \begin{pmatrix} 2 & ? & 1 & 1 \\ ? & 2 & ? & 1 \\ 1 & ? & 2 & ? \\ 1 & 1 & ? & 2 \end{pmatrix}.$$

Find the positive definite completion for which the determinant is maximal. Notice that this completion is not Toeplitz. Next, find the positive definite Toeplitz completion for which the determinant is maximal.

44 Find the representation (1.3.10) for the matrix

$$\begin{pmatrix} 2 & -1 & & & -1 \\ -1 & 2 & -1 & & \\ & \ddots & \ddots & \ddots & \\ & & -1 & 2 & -1 \\ -1 & & & -1 & 2 \end{pmatrix}. \tag{1.6.7}$$

45 Let A be a positive semidefinite Toeplitz matrix. Find a closest (in the Frobenius norm) rank 1 positive semidefinite Toeplitz matrix A_0. When is the solution unique? (Hint: use (1.3.10).) Is the solution unique in the case of the matrix (1.6.7)?

46 Let $A = (A_{j-i})_{i,j=1}^n$ be a positive definite Toeplitz matrix, $0 < p_1 < p_2 < \cdots < p_r < n$ and $\alpha_1, \ldots, \alpha_r \in \mathbb{C}$. Then there exists a unique Toeplitz matrix $F = (F_{j-i})_{i,j=1}^n$ with $F_q = A_q$, $|q| \neq p_1, \ldots, p_r$, and

$$\sum_{j-i=p_k} (F^{-1})_{ij} = \alpha_k \ , \ k = 1, \ldots, r.$$

In case A and $\alpha_1, \ldots, \alpha_r$ are real, the matrix F is also real. (Hint: Consider in Theorem 1.4.3 $B = (B_{ij})_{i,j=1}^n$ with $B_{1p_k} = B_{p_k 1}^* = \alpha_k$, $k = 1, \ldots, r$ and $B_{ij} = 0$ otherwise, and let $\mathcal{W}$ be the set of all Hermitian Toeplitz matrices which have entries equal to 0 on all diagonals except $\pm p_1, \ldots, \pm p_r$.)

47 Recall that $s_1(X) \geq s_2(X) \geq \cdots$ denote the singular values of the matrix X.

(a) Show that for positive definite $n \times n$ matrices A and B it holds that

$$\operatorname{tr}(A-B)(B^{-1}-A^{-1}) \geq \frac{\sum_{j=1}^{n} s_j(A-B)^2}{s_1(A)s_1(B)}.$$

(b) Use (a) to give an alternative proof of Corollary 1.4.4 that a positive definite matrix is uniquely determined by some of its entries and the complementary entries of its inverse (in fact, the statement is slightly stronger than Corollary 1.4.4, as we do not require the given entries in the matrix to include the main diagonal).

48 Prove Theorem 1.4.9.

49 Let $p_j = \overline{p}_{-j}$ be complex numbers and introduce the convex set

$$\mathcal{F} = \left\{ F = (F_{j,k})_{j,k=0}^{n} \geq 0 \ : \ \sum_{l=0}^{n-j} F_{l+j,l} = p_j, j = 0, \ldots, n \right\},$$

and assume that $\mathcal{F} \neq \emptyset$. Let $F_{\text{opt}} \in \mathcal{F}$ be so that $(F_{\text{opt}})_{00} \geq F_{00}$ for all $F \in \mathcal{F}$. Show that F_{opt} is of rank 1. (Hint: Let

$$F_{\text{opt}} = \begin{pmatrix} a & 0_{1\times n} \\ b & F_1 \end{pmatrix} \begin{pmatrix} \overline{a} & b^* \\ 0_{n\times 1} & F_1^* \end{pmatrix}$$

be a Cholesky factorization of F_{opt}. If F_1 has a nonzero upper left entry, argue that

$$F_{\text{opt}} - \begin{pmatrix} 0_{1\times 1} & 0_{1\times n} \\ 0_{n\times 1} & F_1 F_1^* \end{pmatrix} + \begin{pmatrix} F_1 F_1^* & 0_{n\times 1} \\ 0_{1\times n} & 0_{1\times 1} \end{pmatrix}$$

lies in $\mathcal{F}$ but has a larger upper left entry than F_{opt}. Adjust the argument to also include the case when $F_1 \neq 0$ but happens to have a zero upper left entry.)

Remark. Notice that the above argument provides a simple iterative method to find $F \in \mathcal{F}$ whose upper left entry is maximal. In principle one could employ this naive procedure to find a solution to the problem in Example 1.5.2.

1.7 NOTES

Section 1.1

The general results about cones and their duals, as well as those about positive (semi)definite matrices in Section 1.1 can be found in works like [91], [485], [108], [21], [319], and [538]. The proof of the Fejér-Riesz theorem dates back to [213] and can also be found in many contemporary texts (see, e.g., [53] and [510]). In Section 2.4 we present its generalization for operator-valued polynomials.

Section 1.2

Matrices which are inverses of tridiagonal ones were first characterized in [80] and matrices which are the inverses of banded ones in [81]. The theory of positive definite completions started in [114] with the basic 3×3 block matrix case. Next, banded partially positive definite block matrices were considered in [192]. Theorems 2.1.1 and 2.1.4 were first proven there in the finite dimensional case.

In [285] undirected graphs were associated with the patterns of partially positive semidefinite matrices and the importance of chordal graphs in this context was discovered. Proposition 1.2.9 is also taken from [285]. The approach based on cones and duality to study positive semidefinite extensions originates in [452] (see also [315] for the block tridiagonal case). Theorem 1.2.10 was first obtained there. The equality $(\mathrm{PSD}_G)^* = \mathrm{PPSD}_G$ for a chordal graph G was also established there, in a way similar to that in Section 1.2. The extreme rays of PSD_G for nonchordal graphs G were studied in [9], [309], [311], and [518]. Example 1.2.11 is taken from [9]. All notation, definitions, and results in Section 1.2 concerning graph theory follow the book [276].

Section 1.3

The approach in Section 1.3 for characterizing finite subsets of $\mathbb{Z}^d$ which have the extension property was developed in [115] and [503] (see also [510]). The characterization in [503] reduces the problem to the equivalence between $\mathbb{T}\mathrm{Pol}_\Lambda^2 = \mathbb{T}\mathrm{Pol}_S^+$ and $\mathcal{B}_S = \mathcal{A}_\Lambda$. Theorem 1.3.3 was taken from [314], which contains the simplest version of Bochner's theorem, namely the one for $G = \mathbb{Z}$.

Theorem 1.3.6 was originally proved by Carathéodory and Fejér in [120] by using a result in [119] by which the extremal rays of the cone of $n \times n$ positive semidefinite Toeplitz matrices are generated by rank 1 such matrices; see also [281, Section 4.1] for an excellent account. Corollary 1.3.7 may be found in [133].

Theorem 1.3.5 was first obtained in [240] by a different method. The theorem replicates a result by Naimark [17] characterizing the extremal solutions of the Hamburger moment problem. Theorems 1.3.8 through Corollary 1.3.16 were also first proven in [240]. However, we found an inaccuracy in the original statement and proof of Lemma 1.3.15. Still, Theorem 1.3.11

holds, which is the key for the proof of the main result, Corollary 1.3.16. We thank the author of that article for his kind help.

Theorem 1.3.20 comes from [58]. The first proofs of Corollary 1.3.34, Corollary 1.3.40, and Theorem 1.3.18 were obtained in [59]. A reference on the Smith normal forms for matrices with integer entries is [415]. Theorem 1.3.25 is due to [336].

Theorem 1.3.22 was proved for $R(N, N)$, $N \geq 3$, in [115] and [503]. That $R(2, 2)$ and $R(1, 1, 1)$ (see Theorem 1.3.17) do not have the extension property appears in [506] (see also [507]). A theorem by Hilbert [317] states that for every $N \geq 3$, there exist real polynomials $F(s, t)$ of degree $2N$ which are positive for every (s, t), but can not be represented as sums of squares of polynomials. The first explicit example of such polynomial was found in [426] (see Exercise 1.6.27(b)). Other parts from Exercise 1.6.27 came from [447], [398], and [103]. Many other examples appear in [483]. The polynomial in Lemma 1.3.23 and its proof can be found in [90] (see also [510]). Its use makes Theorem 1.3.22 be true for $R(2, 2)$ as well ([510]).

For application of positive trigonometric polynomials to signal processing, see, for instance, [188].

Section 1.4

Theorem 1.4.3 and Corollary 1.4.4 are taken from [69]. Corollary 1.4.4 appears earlier in [525, Theorem 1] as a statement on Gaussian measures with prescribed margins. The modified Newton algorithm used in Example 1.5.1 can be found in [433] (see also [106]). For a semidefinite programming approach, see [26]. Corollaries 1.4.4 and 1.4.5 were first proved in [219] (see also [285]). Exercise 1.6.47, which is based on [219], shows the reasoning there. In [70] an operator-valued version of Corollary 1.4.5 appears. The results in Subsection 1.4.2 are presented for polynomials of a single variable, but they similarly hold for polynomials of several variables as well. For a matrix-valued correspondent of Theorem 1.4.9 see Theorem 2.3.9. The fact that diagonally strictly dominant Hermitian matrices are positive definite follows directly from the Geršgorin disk theorem; see for example [323].

Section 1.5

Semidefinite programming (and the more general area of conic optimization) is an area that has developed developed during the past two decades. It is not our aim to give here a detailed account on this subject. The articles and books [106], [433], [107], [554], [108], [258], and [170] are good sources for information on this topic.

Exercises

Exercise 1.6.9 is based on [286] (parts (c), (e)) and [131] (part (b), (d)); see also [409] and [404] for more results on extreme points of the set of correlation matrices.

Exercise 1.6.10 (a) and (b) are based on [518, Theorem B], which solved a conjecture from [309]. Part (c) of this exercise is based on [309]. As a

separate note, let us remark that the computation of $\beta(G)$ is an NP-complete problem; see [582]. In addition, all graphs for which the extreme rays are generated by matrices of rank at most 2 are characterized in [395].

The answer to Exercise 1.6.11 can be found in [9].

Exercise 1.6.12 is based on a result in [309].

Exercise 1.6.18 is based on [583]. In [316] the problem of finding a closest correlation matrix is considered, motivated by a problem in finance.

Exercise 1.6.23 is based on the proof of Lemma 2.2 in [336].

Exercise 1.6.24 is a result proved by N. Akhiezer and M.G. Krein (see [376]; in [25] the matrix-valued case appears), while Exercises 1.6.42 and 1.6.46 are results from [69].

Exercise 1.6.28 is based on [95]. The Robinson polynomial appears first in the paper [484]; see also [483] for a further discussion.

Exercises 1.6.32 and 1.6.33 come from [58].

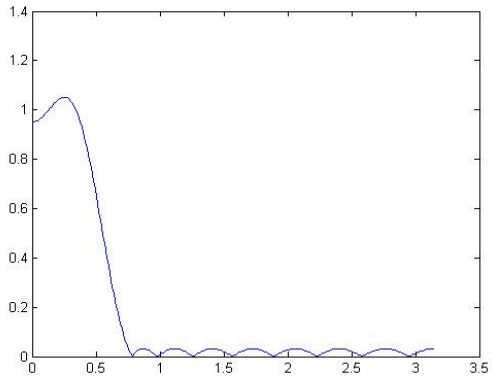
The operations Φ and Ψ from Exercise 1.6.38 are discussed in [330].

Exercise 1.6.43 is based on an example from [412], and the answer for Exercise 1.6.45 may be found in [533]. The matrix (1.6.7), as well as minor variations, is discussed in [531].

Additional notes

The standard approach to the *frequency-domain design* of one-dimensional *recursive digital filters* begins by finding the *squared magnitude response* of a filter that best approximates the desired response. *Spectral factorization* is then performed to find the (unique) minimum-phase filter that has that magnitude response, which is guaranteed to be stable. The design problem, embodied by the determination of the squared magnitude function, and the implementation problem, embodied by the spectral factorization, are decoupled.

Consider a class of systems whose input u and output y satisfy a *linear*



constant coefficient difference equation of the form

$$\sum_{k=0}^{n} a_k y_{t-k} = \sum_{k=0}^{n} b_k u_k. \tag{1.7.1}$$

The frequency response, that is, the system transfer function evaluated on the unit circle, has the form

$$f(z) = \frac{\sum\limits_{k=0}^{n} a_k z^k}{\sum\limits_{k=0}^{n} b_k z^k}.$$

In order to design a *low pass filter*, the constraints we have to satisfy are best written using the squared magnitude of the filter frequency response $|f(z)|^2$. In designing a lowpass filter one would like $|f(z)|^2 \approx 1$ when $z \in \{e^{it} : t \in [0, s]\}$ for some $0 < s < \pi$, and $|f(z)|^2 \approx 0$ when $z \in \{e^{it} : t \in [s, \pi]\}$. A typical squared magnitude response of a low pass filter looks like the graph above (see for example [581] or [244]).

Spectral factorization is now performed to find the coefficients a_k, b_k of the minimum-phase filter that has the desired magnitude response, which is guaranteed to be stable. The coefficients may then be used to implement the filter. General references on the subject include [347], [584], [27], [561], [469], [446], [589], [333], [107], and [432].