22 Tools

This chapter collects some mathematical tools used in this book: (direct) convex separation results in Sects. 22.2 and 22.6; Lemmas of the Alternative, in particular Farkas' Lemma in Sect. 22.3; the Linear Duality Theorem in Sect. 22.4; the Brouwer and Kakutani Fixed Point Theorems in Sect. 22.5; the Krein–Milman Theorem and the Birkhoff–von Neumann Theorem in Sect. 22.6; and the Max Flow Min Cut Theorem of Ford and Fulkerson in Sect. 22.7.

22.1 Some Definitions

A subset $Z \subseteq \mathbb{R}^n$ is *convex* if with any two points $\mathbf{x}, \mathbf{y} \in Z$, also the line segment connecting $\mathbf{x}$ and $\mathbf{y}$ is contained in Z. Formally:

$$\forall\, \mathbf{x}, \mathbf{y} \in Z\ \forall\, 0 \le \lambda \le 1 : \lambda \mathbf{x} + (1-\lambda)\mathbf{y} \in Z\ .$$

If Z is a closed set[1] then for convexity it is sufficient to check this condition for $\lambda = 1/2$ (see Problem 22.1). It is easy to see that a set $Z \subseteq \mathbb{R}^n$ is convex if and only if $\sum_{j=1}^k \lambda_j \mathbf{x}^j \in Z$ for all $\mathbf{x}^1, \ldots, \mathbf{x}^k \in Z$ and all nonnegative $\lambda_1, \ldots, \lambda_k \in \mathbb{R}$ with $\sum_{j=1}^k \lambda_j = 1$. Such a sum $\sum_{j=1}^k \lambda_j \mathbf{x}^j$ is called a *convex combination* of the $\mathbf{x}^j$. For an arbitrary subset $D \subseteq \mathbb{R}^n$, the *convex hull of D* is the set of all convex combinations of elements of D or, equivalently, the smallest (with respect to set inclusion) convex subset of $\mathbb{R}^n$ containing D.

For vectors $\mathbf{x} = (x_1, \ldots, x_n)$, $\mathbf{y} = (y_1, \ldots, y_n) \in \mathbb{R}^n$,

$$\mathbf{x} \cdot \mathbf{y} := \sum_{i=1}^n x_i y_i$$

[1] A set $Z \subseteq \mathbb{R}^n$ is *closed* if it contains the limit of every converging sequence in Z.

H. Peters, *Game Theory*, Springer Texts in Business and Economics,
DOI 10.1007/978-3-662-46950-7_22

denotes the *inner product* of $\mathbf{x}$ and $\mathbf{y}$, and

$$||\mathbf{x}-\mathbf{y}|| := \sqrt{\sum_{i=1}^{n}(x_i-y_i)^2}$$

is the *Euclidean distance* between $\mathbf{x}$ and $\mathbf{y}$. A set $C \subseteq \mathbb{R}^n$ is a (convex) *cone* if, with each $\mathbf{x}, \mathbf{y} \in C$ and $\lambda \in \mathbb{R}$, $\lambda \geq 0$, also $\lambda\mathbf{x} \in C$ and $\mathbf{x}+\mathbf{y} \in C$.

22.2 A Separation Theorem

In this section we derive the simplest version of a separation result, namely separating a point from a convex set.

Theorem 22.1 *Let $Z \subseteq \mathbb{R}^n$ be a closed convex set and let $\mathbf{x} \in \mathbb{R}^n \setminus Z$. Then there is a $\mathbf{y} \in \mathbb{R}^n$ with $\mathbf{y}\cdot\mathbf{z} > \mathbf{y}\cdot\mathbf{x}$ for every $\mathbf{z} \in Z$.*

Thus, this theorem states the geometrically obvious fact that a closed convex set and a point not in that set can be *separated* by a hyperplane (with normal $\mathbf{y}$).

Proof of Theorem 22.1 Let $\mathbf{z}' \in Z$ such that $0 < ||\mathbf{x}-\mathbf{z}'|| \leq ||\mathbf{x}-\mathbf{z}||$ for all $\mathbf{z} \in Z$. Such a $\mathbf{z}'$ exists by the Theorem of Weierstrass, since the Euclidean distance from $\mathbf{x}$ is a continuous function on the set Z, and for the minimum of $\mathbf{z} \to ||\mathbf{x}-\mathbf{z}||$ on Z attention can be restricted to a compact (i.e., bounded and closed) subset of Z. Let $\mathbf{y} = \mathbf{z}' - \mathbf{x}$. Let $\mathbf{z} \in Z$. For any α, $0 \leq \alpha \leq 1$, convexity of Z implies $\mathbf{z}' + \alpha(\mathbf{z}-\mathbf{z}') \in Z$, and thus

$$||\mathbf{z}' + \alpha(\mathbf{z}-\mathbf{z}') - \mathbf{x}||^2 \geq ||\mathbf{z}'-\mathbf{x}||^2 \,.$$

Hence,

$$2\alpha(\mathbf{z}'-\mathbf{x})\cdot(\mathbf{z}-\mathbf{z}') + \alpha^2||\mathbf{z}-\mathbf{z}'||^2 \geq 0 \,.$$

Thus, letting $\alpha \downarrow 0$, it follows that $(\mathbf{z}'-\mathbf{x})\cdot(\mathbf{z}-\mathbf{z}') \geq 0$. From this, $(\mathbf{z}'-\mathbf{x})\cdot\mathbf{z} \geq (\mathbf{z}'-\mathbf{x})\cdot\mathbf{z}' = (\mathbf{z}'-\mathbf{x})\cdot\mathbf{x} + (\mathbf{z}'-\mathbf{x})\cdot(\mathbf{z}'-\mathbf{x}) > (\mathbf{z}'-\mathbf{x})\cdot\mathbf{x}$.

Because $\mathbf{z}$ was arbitrary, it follows that $\mathbf{y}\cdot\mathbf{z} > \mathbf{y}\cdot\mathbf{x}$ for every $\mathbf{z} \in Z$. ■

Remark 22.2 A consequence of Theorem 22.1 is that there are real numbers α and β satisfying $\mathbf{y}\cdot\mathbf{z} > \alpha$ and $\mathbf{y}\cdot\mathbf{x} < \alpha$, and $\mathbf{y}\cdot\mathbf{z} > \beta$ and $\mathbf{y}\cdot\mathbf{x} = \beta$, for all $\mathbf{z} \in Z$ (notations as in the theorem). The last assertion is trivial. For the first assertion, note that in the proof of the theorem we have $\mathbf{y}\cdot\mathbf{z} \geq \mathbf{y}\cdot\mathbf{z}'$ for all $\mathbf{z} \in Z$, so $\mathbf{y}\cdot\mathbf{z}'$ is a lower bound for $\mathbf{y}\cdot\mathbf{z}$. Then take, for instance, $\alpha = \frac{1}{2}(\mathbf{y}\cdot\mathbf{z}' + \mathbf{y}\cdot\mathbf{x})$.

22.3 Lemmas of the Alternative

Theorem 22.1 can be used to derive several *lemmas of the alternative*. These lemmas have in common that they describe two systems of linear inequalities and equations, exactly one of which has a solution.

Lemma 22.3 (Theorem of the Alternative for Matrices) *Let A be an $m \times n$ matrix. Exactly one of the following two statements is true.*

(a) *There are $\mathbf{y} \in \mathbb{R}^n$ and $\mathbf{z} \in \mathbb{R}^m$ with $(\mathbf{y}, \mathbf{z}) \geq \mathbf{0}$, $(\mathbf{y}, \mathbf{z}) \neq \mathbf{0}$ and $A\mathbf{y} + \mathbf{z} = 0$.*
(b) *There is an $\mathbf{x} \in \mathbb{R}^m$ with $\mathbf{x} > \mathbf{0}$ and $\mathbf{x}A > \mathbf{0}$.*

Proof We leave it to the reader to prove that at most one of the systems in (a) and (b) has a solution (Problem 22.2).

Now suppose that (a) is not true. It is sufficient to prove that the system in (b) must have a solution. Observe that (a) implies that $\mathbf{0}$ is a convex combination of the columns of A and the set $\{\mathbf{e}^j \in \mathbb{R}^m \mid j = 1, \ldots, m\}$. This follows from dividing both sides of the equation $A\mathbf{y} + \mathbf{z} = 0$ by $\sum_{j=1}^n y_j + \sum_{i=1}^m z_i$. Hence, the assumption that (a) is not true means that $\mathbf{0} \notin Z$, where $Z \subseteq \mathbb{R}^m$ is the convex hull of the columns of A and the set $\{\mathbf{e}^j \in \mathbb{R}^m \mid j = 1, \ldots, m\}$. By Theorem 22.1 and Remark 22.2 there is an $\mathbf{x} \in \mathbb{R}^m$ and a number $\beta \in \mathbb{R}$ such that $\mathbf{x} \cdot \mathbf{p} > \beta$ for all $p \in Z$ and $\mathbf{x} \cdot \mathbf{0} = \beta$. Hence, $\beta = 0$ and, in particular, $\mathbf{x}A > \mathbf{0}$ and $\mathbf{x} > \mathbf{0}$ since the columns of A and all $\mathbf{e}^j$ for $j = 1, \ldots, m$ are elements of Z. Thus, (b) is true. ■

Another lemma of the alternative is Farkas' Lemma below. In its proof we use the following result.

Lemma 22.4 *Let A be an $m \times n$ matrix and let*

$$Z = \{\mathbf{z} \in \mathbb{R}^n \mid \text{ there exists an } \mathbf{x} \in \mathbb{R}^m,\ \mathbf{x} \geq \mathbf{0} \text{ with } \mathbf{z} = \mathbf{x}A\}.$$

Then Z is closed.

Proof

(a) Suppose that the rank of the matrix A, $r(A)$, is equal to m. Then also $r(AA^T) = m$, where A^T is the transpose of A (Problem 22.3). Therefore, AA^T is invertible. Let $(\mathbf{z}^n)_{n \in \mathbb{N}}$ be a sequence in Z converging to $\mathbf{z} \in \mathbb{R}^n$. Let $\mathbf{z}^n = \mathbf{x}^n A$, $\mathbf{x}^n \geq \mathbf{0}$, for all $n \in \mathbb{N}$. Then $\mathbf{x}^n = \mathbf{x}^n AA^T(AA^T)^{-1}$ for every n, hence $\mathbf{x}^n A \to \mathbf{z}$ implies $\mathbf{x}^n \to \mathbf{z}A^T(AA^T)^{-1} =: \mathbf{x}$, and in particular $\mathbf{x} \geq \mathbf{0}$. Thus, since $\mathbf{x}^n A \to \mathbf{x}A$, we obtain $\mathbf{z} = \mathbf{x}A$, so that $\mathbf{z} \in Z$.
(b) Let $b \in Z \setminus \{\mathbf{0}\}$ and choose $\mathbf{x} \in \mathbb{R}^m$, $\mathbf{x} \geq \mathbf{0}$, with $\mathbf{b} = \mathbf{x}A$ such that $|S|$ is maximal, where $S := \{i \in \{1, \ldots, m\} \mid x_i > 0\}$. We show that the rows of A with numbers in S are linearly independent. If not, then there is a $\boldsymbol{\mu} \in \mathbb{R}^m$ with $\boldsymbol{\mu}A = \mathbf{0}$,

$\mu_j \neq 0$ for some $j \in S$, and $\mu_i = 0$ for all $i \notin S$. Then $(\mathbf{x} - t\boldsymbol{\mu})A = \mathbf{x}A = \mathbf{b}$ for every $t \in \mathbb{R}$. Choose $\hat{t}$ in such a way that $x_j - \hat{t}\mu_j \geq 0$ for all $j \in S$ and $x_j - \hat{t}\mu_j = 0$ for some $j \in S$. Then $\mathbf{b} = (\mathbf{x} - \hat{t}\boldsymbol{\mu})A$, $\mathbf{x} - \hat{t}\boldsymbol{\mu} \geq \mathbf{0}$, and $|\{i \in \{1, \ldots, m\} \mid x_i - t\mu_i > 0\}| \leq |S| - 1$, a contradiction.

(c) In view of part (b) of the proof we can write

$$Z = \bigcup_B \{\mathbf{x}B \mid B \text{ a } k \times n \text{ submatrix of } A,\ r(B) = k \leq r(A),\ \mathbf{0} \leq \mathbf{x} \in \mathbb{R}^k\} .$$

By part (a), each of the sets at the right hand side of this equation is closed. Hence, Z is the union of finitely many closed sets, and therefore is closed itself. ∎

Lemma 22.5 (Farkas' Lemma) *Let A be an $m \times n$ matrix and $\mathbf{b} \in \mathbb{R}^n$. Exactly one of the following two statements is true.*

(a) *There is an $\mathbf{x} \in \mathbb{R}^m$ with $\mathbf{x} \geq \mathbf{0}$ and $\mathbf{x}A = \mathbf{b}$.*
(b) *There is a $\mathbf{y} \in \mathbb{R}^n$ with $A\mathbf{y} \geq \mathbf{0}$ and $\mathbf{b} \cdot \mathbf{y} < \mathbf{0}$.*

Proof We leave it to the reader to show that at most one of the two systems in (a) and (b) can have a solution (Problem 22.4). Assume that the system in (a) does not have a solution. It is sufficient to prove that the system in (b) must have a solution.

The assumption that the system in (a) does not have a solution is equivalent to the statement $\mathbf{b} \notin Z$ where

$$Z = \{\mathbf{z} \in \mathbb{R}^n \mid \text{ there exists an } \mathbf{x} \in \mathbb{R}^m, \mathbf{x} \geq \mathbf{0} \text{ with } \mathbf{z} = \mathbf{x}A\}.$$

Clearly, the set Z is convex, and it is closed by Lemma 22.4. By Theorem 22.1 and Remark 22.2 it follows that there is a $\mathbf{y} \in \mathbb{R}^n$ and an $\alpha \in \mathbb{R}$ with $\mathbf{y} \cdot \mathbf{b} < \alpha$ and $\mathbf{y} \cdot \mathbf{z} > \alpha$ for all $\mathbf{z} \in Z$. Because $\mathbf{0} \in Z$ it follows that $\alpha < \mathbf{y} \cdot \mathbf{0} = 0$, hence $\mathbf{y} \cdot \mathbf{b} < \alpha < 0$. To prove that the system in (b) has a solution, it is sufficient to prove that $A\mathbf{y} \geq \mathbf{0}$. Suppose not, i.e., there is an i with $(A\mathbf{y})_i < 0$. Then $\mathbf{e}^i A\mathbf{y} < 0$, so $(M\mathbf{e}^i)A\mathbf{y} \to -\infty$ as $\mathbb{R} \ni M \to \infty$. Observe, however, that $(M\mathbf{e}^i)A \in Z$ for every $M > 0$, so that $(M\mathbf{e}^i)A\mathbf{y} > \alpha$ for every such M. This contradiction completes the proof of the lemma. ∎

These lemmas can be interpreted geometrically. We show this for Farkas' Lemma in Fig. 22.1. Consider the row vectors $\mathbf{r}_i$ of A as points in $\mathbb{R}^n$. The set of all nonnegative linear combinations of the $\mathbf{r}_i$ forms a cone C. The statement that the system in (i) in Lemma 22.5 has no nonnegative solution means that the vector $\mathbf{b}$ does not lie in C. In this case, the lemma asserts the existence of a vector $\mathbf{y}$ which makes an obtuse angle with $\mathbf{b}$ and a nonobtuse angle with each of the vectors $\mathbf{r}_i$. This means that the hyperplane L orthogonal to $\mathbf{y}$ has the cone C on one side and the point $\mathbf{b}$ on the other.

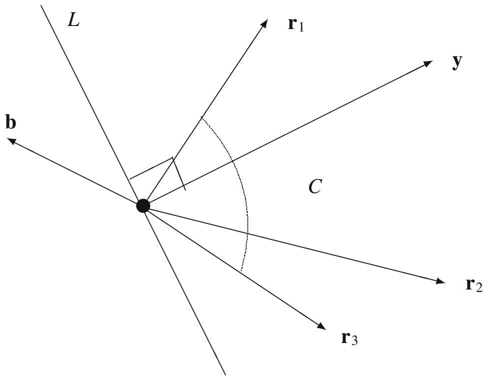

Fig. 22.1 Geometric interpretation of Farkas' Lemma

22.4 The Duality Theorem of Linear Programming

In this section we prove the following theorem.

Theorem 22.6 (Duality Theorem of Linear Programming) *Let A be an $n \times p$ matrix, $\mathbf{b} \in \mathbb{R}^p$, and $\mathbf{c} \in \mathbb{R}^n$. Suppose $V := \{\mathbf{x} \in \mathbb{R}^n \mid \mathbf{x}A \geq \mathbf{b}, \mathbf{x} \geq \mathbf{0}\} \neq \emptyset$ and $W := \{\mathbf{y} \in \mathbb{R}^p \mid A\mathbf{y} \leq \mathbf{c}, \mathbf{y} \geq \mathbf{0}\} \neq \emptyset$. Then* $\min\{\mathbf{x}\cdot\mathbf{c} \mid \mathbf{x} \in V\} = \max\{\mathbf{b}\cdot\mathbf{y} \mid \mathbf{y} \in W\}$.

To prove this theorem, we first prove the following variant of Farkas' Lemma.

Lemma 22.7 *Let A be an $m \times n$ matrix and $\mathbf{b} \in \mathbb{R}^n$. Exactly one of the following two statements is true.*

(a) *There is an $\mathbf{x} \in \mathbb{R}^m$ with $\mathbf{x}A \leq \mathbf{b}$ and $\mathbf{x} \geq \mathbf{0}$.*
(b) *There is a $\mathbf{y} \in \mathbb{R}^n$ with $A\mathbf{y} \geq \mathbf{0}$, $\mathbf{b} \cdot \mathbf{y} < 0$, and $\mathbf{y} \geq \mathbf{0}$.*

Proof Problem 22.5. ■

The following two lemmas are further preparations for the proof of the Duality Theorem.

Lemma 22.8 *Let $\mathbf{x} \in V$ and $\mathbf{y} \in W$ (cf. Theorem 22.6). Then $\mathbf{x} \cdot \mathbf{c} \geq \mathbf{b} \cdot \mathbf{y}$.*

Proof $\mathbf{x} \cdot \mathbf{c} \geq \mathbf{x}A\mathbf{y} \geq \mathbf{b} \cdot \mathbf{y}$. ■

Lemma 22.9 *Let $\hat{\mathbf{x}} \in V$, $\hat{\mathbf{y}} \in W$ with $\hat{\mathbf{x}} \cdot \mathbf{c} = \mathbf{b} \cdot \hat{\mathbf{y}}$. Then $\hat{\mathbf{x}} \cdot \mathbf{c} = \min\{\mathbf{x} \cdot \mathbf{c} \mid \mathbf{x} \in V\}$ and $\mathbf{b} \cdot \hat{\mathbf{y}} = \max\{\mathbf{b} \cdot \mathbf{y} \mid \mathbf{y} \in W\}$.*

Proof By Lemma 22.8, for every $\mathbf{x} \in V$: $\mathbf{x} \cdot \mathbf{c} \geq \mathbf{b} \cdot \hat{\mathbf{y}} = \hat{\mathbf{x}} \cdot \mathbf{c}$. Similarly, $\mathbf{b} \cdot \mathbf{y} \leq \hat{\mathbf{x}} \cdot \mathbf{c} = \mathbf{b} \cdot \hat{\mathbf{y}}$ for every $\mathbf{y} \in W$. ■

Proof of Theorem 22.6 In view of Lemmas 22.8 and 22.9, it is sufficient to show the existence of $\hat{\mathbf{x}} \in V$ and $\hat{\mathbf{y}} \in W$ with $\hat{\mathbf{x}} \cdot \mathbf{c} \leq \mathbf{b} \cdot \hat{\mathbf{y}}$. So it is sufficient to show that the system

$$(\mathbf{x}, \mathbf{y}) \begin{pmatrix} -A & \mathbf{0} & \mathbf{c} \\ \mathbf{0} & A^T & -\mathbf{b} \end{pmatrix} \leq (-\mathbf{b}, \mathbf{c}, \mathbf{0}), \;\; \mathbf{x} \geq \mathbf{0}, \;\; \mathbf{y} \geq \mathbf{0}$$

has a solution. Suppose this is not the case. By Lemma 22.7, there exists a vector $(\mathbf{z}, \mathbf{w}, t) \in \mathbb{R}^p \times \mathbb{R}^n \times \mathbb{R}$ with

$$\begin{pmatrix} -A & \mathbf{0} & \mathbf{c} \\ \mathbf{0} & A^T & -\mathbf{b} \end{pmatrix} \begin{pmatrix} \mathbf{z} \\ \mathbf{w} \\ t \end{pmatrix} \geq \mathbf{0}, \;\; (-\mathbf{b}, \mathbf{c}, 0) \cdot (\mathbf{z}, \mathbf{w}, t) < \mathbf{0}, \;\; \mathbf{z} \geq \mathbf{0}, \;\; \mathbf{w} \geq \mathbf{0}, \;\; t \geq 0 \,.$$

Hence

$$A\mathbf{z} \leq t\mathbf{c} \tag{22.1}$$

$$\mathbf{w}A \geq t\mathbf{b} \tag{22.2}$$

$$\mathbf{c} \cdot \mathbf{w} < \mathbf{b} \cdot \mathbf{z} \,. \tag{22.3}$$

If $t = 0$, then $A\mathbf{z} \leq \mathbf{0}$ and $\mathbf{w}A \geq \mathbf{0}$, hence, for $\mathbf{x} \in V$ and $\mathbf{y} \in W$:

$$\mathbf{b} \cdot \mathbf{z} \leq \mathbf{x}A\mathbf{z} \leq 0 \leq \mathbf{w}A\mathbf{y} \leq \mathbf{w} \cdot \mathbf{c}$$

contradicting (22.3). If $t > 0$, then by (22.1) and (22.2), $t^{-1}\mathbf{z} \in W$ and $t^{-1}\mathbf{w} \in V$. By (22.3), $\mathbf{b} \cdot (t^{-1}\mathbf{z}) > (t^{-1}\mathbf{w}) \cdot \mathbf{c}$, which contradicts Lemma 22.8. Hence, the first system above must have a solution. ■

22.5 Some Fixed Point Theorems

Let $Z \subseteq \mathbb{R}^n$ be a nonempty convex and compact[2] set. Let $f : Z \to Z$ be a continuous function. A point $\mathbf{x}^* \in Z$ is a *fixed point* of f if $f(\mathbf{x}^*) = \mathbf{x}^*$.

If $n = 1$, then Z is a closed interval of the form $[a, b] \subseteq \mathbb{R}$, and then it is clear (by drawing a picture) that f must have a fixed point: formally, this is a straightforward implication of the intermediate-value theorem.

More generally, the following result holds.

[2]A set $Z \subseteq \mathbb{R}^n$ is *compact* if it is closed and bounded. A set $Z \subseteq \mathbb{R}^n$ is bounded if there is an $M > 0$ such that $||\mathbf{x}|| < M$ for all $\mathbf{x} \in Z$.

Theorem 22.10 (Brouwer Fixed Point Theorem) *Let $Z \subseteq \mathbb{R}^n$ be a nonempty compact and convex set and let $f : Z \to Z$ be a continuous function. Then f has a fixed point.*

A generalization of Brouwer's fixed point theorem is Kakutani's fixed point theorem. Let $F : Z \to Z$ be a *correspondence*, i.e., $F(\mathbf{x})$ is a nonempty subset of Z for every $\mathbf{x} \in Z$. Call F *convex-valued* if $F(\mathbf{x})$ is a convex set for every $\mathbf{x} \in Z$. Call F *upper semi-continuous* if the following holds: for every sequence $(\mathbf{x}^k)_{k\in\mathbb{N}}$ in Z converging to $\mathbf{x} \in Z$ and for every sequence $(\mathbf{y}^k)_{k\in\mathbb{N}}$ in Z converging to $\mathbf{y} \in Z$, if $\mathbf{y}^k \in F(\mathbf{x}^k)$ for every $k \in \mathbb{N}$, then $\mathbf{y} \in F(\mathbf{x})$. A point $\mathbf{x}^* \in Z$ is a *fixed point* of Z if $\mathbf{x}^* \in F(\mathbf{x}^*)$.

Theorem 22.11 (Kakutani Fixed Point Theorem) *Let $Z \subseteq \mathbb{R}^n$ be a nonempty compact and convex set and let $F : Z \to Z$ be an upper semi-continuous and convex-valued correspondence. Then F has a fixed point.*

22.6 The Birkhoff–von Neumann Theorem

Let C be a convex set in a linear space V. An element $e \in C$ is called an *extreme point* of C if for all $x, y \in C$ with $e = \frac{1}{2}(x + y)$ it holds that $x = y$ $(= e)$. By $\text{ext}(C)$ the set of extreme points of C is denoted. See Problem 22.6 for alternative characterizations of extreme points.

An $n \times n$-matrix D is called *doubly stochastic* if $0 \leq d_{ij} \leq 1$ for all $i, j = 1, \ldots, n$, $\sum_{j=1}^{n} d_{ij} = 1$ for all i, and $\sum_{i=1}^{n} d_{ij} = 1$ for all j. If moreover $d_{ij} \in \{0, 1\}$ for all $i, j = 1, \ldots, n$, then D is called a *permutation matrix*. Let $D_{n\times n}$ denote the set of all $n \times n$ doubly stochastic matrices, and let $P_{n\times n}$ denote the set of all $n \times n$ permutation matrices. Note that $D_{n\times n}$ is a convex compact set, and that $P_{n\times n}$ is a finite subset of $D_{n\times n}$. The following theorem gives the exact relation.

Theorem 22.12 (Birkhoff–von Neumann)

(a) $\text{ext}(D_{n\times n}) = P_{n\times n}$
(b) $D_{n\times n} = \text{conv}(P_{n\times n})$.

Part (b) of Theorem 22.12 follows from the Theorem of Krein–Milman (Theorem 22.14 below). In the proof of the latter theorem the dimension of a subset of a linear space V plays a role. A subset of V of the form $a + L$ where $a \in V$ and L is a linear subspace of V, is called an *affine subspace*. Check that a subset A of V is affine if, and only if, with any two different elements x and y of A, also the straight line through x and y is contained in A (Problem 22.7). For an affine subspace $a + L$ of V the *dimension* is defined to be the dimension of the linear subspace L. For an arbitrary subset A of V, its *dimension* $\dim(A)$ is defined to be the dimension of the smallest affine subspace of V containing the set A.

The following separation lemma is used in the proof of the Theorem of Krein–Milman. By $\text{int}(C)$ and $\text{clo}(C)$ we denote the (topological) interior and closure of the set $C \subseteq \mathbb{R}^n$, respectively.

Lemma 22.13 *Let C be a nonempty convex subset of $\mathbb{R}^n$ and $\mathbf{a} \in \mathbb{R}^n \setminus \text{int}(C)$. Then there exists a $\mathbf{p} \in \mathbb{R}^n \setminus \{\mathbf{0}\}$ with $\mathbf{p} \cdot \mathbf{a} \leq \mathbf{p} \cdot \mathbf{c}$ for every $\mathbf{c} \in C$.*

Proof We distinguish two cases: (a) $\mathbf{a} \notin \text{clo}(C)$ and (b) $\mathbf{a} \in \text{clo}(C)$.

(a) Suppose $\mathbf{a} \notin \text{clo}(C)$. Then the result follows from Theorem 22.1, with $\text{clo}(C)$ in the role of the set Z there.
(b) Suppose $\mathbf{a} \in \text{clo}(C)$. Because $\mathbf{a} \notin \text{int}(C)$, there is a sequence $\mathbf{a}^1, \mathbf{a}^2, \ldots \in \mathbb{R}^n \setminus \text{clo}(C)$ converging to $\mathbf{a}$. By Theorem 22.1 again, for each k there is a $\mathbf{p}^k \in \mathbb{R}^n \setminus \{\mathbf{0}\}$ with $\mathbf{p}^k \cdot \mathbf{a}^k \leq \mathbf{p}^k \cdot \mathbf{c}$ for all $\mathbf{c} \in \text{clo}(C)$, and we can take these vectors $\mathbf{p}^k$ such that $||\mathbf{p}^k|| = 1$ for every k ($|| \cdot ||$ denotes the Euclidean norm). Because $\{\mathbf{x} \in \mathbb{R}^n \mid ||\mathbf{x}|| = 1\}$ is a compact set, there exists a converging subsequence $\mathbf{p}^{k(1)}, \mathbf{p}^{k(2)}, \ldots$ of $\mathbf{p}^1, \mathbf{p}^2, \ldots$ with limit, say, $\hat{\mathbf{p}}$. Then $\hat{\mathbf{p}} \cdot \mathbf{a} = \lim_{\ell \to \infty} \mathbf{p}^{k(\ell)} \cdot \mathbf{a}^{k(\ell)} \leq \lim_{\ell \to \infty} \mathbf{p}^{k(\ell)} \cdot \mathbf{c} = \hat{\mathbf{p}} \cdot \mathbf{c}$ for all $\mathbf{c} \in \text{clo}(C)$. ■

Theorem 22.14 (Krein–Milman) *Let C be a nonempty compact and convex subset of $\mathbb{R}^n$. Then $\text{ext}(C) \neq \emptyset$ and $C = \text{conv}(\text{ext}(C))$.*

Proof

(1) Because C is compact and $\mathbf{x} \mapsto ||\mathbf{x}||$ (where $|| \cdot ||$ denotes the Euclidean norm) is continuous, there exists by the Theorem of Weierstrass an $\mathbf{e} \in C$ with $||\mathbf{e}|| = \max_{\mathbf{x} \in C} ||\mathbf{x}||$. Then $\mathbf{e} \in \text{ext}(C)$, which can be proved as follows. Suppose that $\mathbf{e} = \frac{1}{2}(\mathbf{x}^1 + \mathbf{x}^2)$ for some $\mathbf{x}^1, \mathbf{x}^2 \in C$. Then

$$||\mathbf{e}|| = ||\frac{1}{2}(\mathbf{x}^1 + \mathbf{x}^2)|| \leq \frac{1}{2}||\mathbf{x}^1|| + \frac{1}{2}||\mathbf{x}^2|| \leq \frac{1}{2}||\mathbf{e}|| + \frac{1}{2}||\mathbf{e}||$$

implies $||\mathbf{x}^1|| = ||\mathbf{x}^2|| = ||\frac{1}{2}(\mathbf{x}^1 + \mathbf{x}^2)||$. By definition of the Euclidean norm this is only possible if $\mathbf{x}^1 = \mathbf{x}^2 = \mathbf{e}$. This shows $\mathbf{e} \in \text{ext}(C)$. Hence, $\text{ext}(C) \neq \emptyset$.
(2) The second statement in the theorem will be proved by induction on $\dim(C)$.
   (a) If $\dim(C) = 0$, then $C = \{\mathbf{a}\}$ for some $\mathbf{a} \in \mathbb{R}^n$, so $\text{ext}(C) = \{\mathbf{a}\}$ and $\text{conv}(\text{ext}(C)) = \{\mathbf{a}\} = C$.
   (b) Let $k \in \mathbb{N}$, and suppose that $\text{conv}(\text{ext}(D)) = D$ for every nonempty compact and convex subset D of $\mathbb{R}^n$ with $\dim(D) < k$. Let C be a k-dimensional compact convex subset of $\mathbb{R}^n$. Obviously, $\text{conv}(\text{ext}(C)) \subseteq C$. So to prove is still: $C \subseteq \text{conv}(\text{ext}(C))$. Without loss of generality assume $\mathbf{0} \in C$ (otherwise, shift the whole set C). Let W be the smallest affine (hence, linear) subset of $\mathbb{R}^n$ containing C. Hence, $\dim(W) = k$. From part (1) of the proof there is an $\mathbf{e} \in \text{ext}(C)$. Let $\mathbf{x} \in C$. If $\mathbf{x} = \mathbf{e}$ then $\mathbf{x} \in \text{conv}(\text{ext}(C))$. If $\mathbf{x} \neq \mathbf{e}$ then the intersection of the straight line through $\mathbf{x}$ and $\mathbf{e}$ with C is a line segment of which one of the endpoints is $\mathbf{e}$. Let $\mathbf{b}$ be the other endpoint.

Then $\mathbf{b}$ is a boundary point of C. Then, by Lemma 22.13, there is a linear function $f : W \to \mathbb{R}$ with $f(\mathbf{b}) = \min\{f(\mathbf{c}) \mid \mathbf{c} \in C\}$ and $f \neq 0$ (check this). Let $D := \{\mathbf{y} \in C \mid f(\mathbf{y}) = f(\mathbf{b})\}$. Then D is a compact and convex subset of C. Because $f \neq 0$ it follows that $\dim(D) < k$. By the induction hypothesis, $D = \text{conv}(\text{ext}(D))$. Also, $\text{ext}(D) \subseteq \text{ext}(C)$, see Problem 22.8. Hence, $\mathbf{b} \in D = \text{conv}(\text{ext}(D)) \subseteq \text{conv}(\text{ext}(C))$. Further, $\mathbf{e} \in \text{ext}(C)$. Because $\mathbf{x} \in \text{conv}\{\mathbf{b}, \mathbf{e}\}$ it follows that $\mathbf{x} \in \text{conv}(\text{ext}(C))$. So $C \subseteq \text{conv}(\text{ext}(C))$. ■

Proof of Theorem 22.12 Because $D_{n\times n}$ is compact and convex, part (b) follows from part (a) and Theorem 22.14. So only (a) still has to be proved.

(1) We first prove that $P_{n\times n} \subseteq \text{ext}(D_{n\times n})$. Let $P = [p_{ij}]_{i,j=1}^n$ be a permutation matrix with $P = \frac{1}{2}(A+B)$ for some $A, B \in D_{n\times n}$. Then $p_{ij} = \frac{1}{2}(a_{ij}+b_{ij})$ and $p_{ij} \in \{0, 1\}$ for all $i, j \in \{1, 2, \ldots, n\}$. If $p_{ij} = 0$ then $a_{ij} = b_{ij} = 0$ because $a_{ij}, b_{ij} \geq 0$. If $p_{ij} = 1$ then $a_{ij} = b_{ij} = 1$ because $a_{ij}, b_{ij} \leq 1$. Hence, $A = B$, so that $P \in \text{ext}(D_{n\times n})$.

(2) Let now $D = [d_{ij}] \in D_{n\times n}$ such that D is not a permutation matrix. The proof is complete if we show that D is not an extreme point. For this, it is sufficient to show that there exists an $n \times n$-matrix $C \neq [0]$ with
   (i) $c_{ij} = 0$ whenever $d_{ij} = 0$ or $d_{ij} = 1$,
   (ii) $\sum_{i=1}^n c_{ij} = 0$ for all $j \in \{1, 2, \ldots, n\}$ with $d_{ij} \neq 1$ for every i,
   (iii) $\sum_{j=1}^n c_{ij} = 0$ for all $i \in \{1, 2, \ldots, n\}$ with $d_{ij} \neq 1$ for every j.
   For in that case, for $\varepsilon > 0$ sufficiently small, the matrices $D + \varepsilon C$ and $D - \varepsilon C$ are two different doubly stochastic matrices with $D = \frac{1}{2}((D+\varepsilon C) + (D-\varepsilon C))$, implying that $D \notin \text{ext}(D_{n\times n})$.

   We are left to construct C. In order to satisfy (i), for those rows or columns of D that contain a 1 the corresponding rows or columns of C contain only zeros. Suppose there are k rows (and hence columns) of D that do not contain a 1. Because D is not a permutation matrix, $2 \leq k \leq n$. In these k rows there are at least $2k$ elements unequal to 0 and 1. The corresponding $2k$ or more elements of C are to be chosen such that they satisfy the system of $2k$ homogeneous linear equations described in (ii) and (iii). Without loss of generality assume that these equations correspond to the first k rows and the first k columns. Then, if $\sum_{j=1}^k c_{ij} = 0$ for all $i \in \{1, \ldots, k-1\}$ and $\sum_{i=1}^k c_{ij} = 0$ for all $j \in \{1, \ldots, k\}$, we have $\sum_{j=1}^k c_{kj} = 0$ as well, so that this last equation is redundant. Thus, we have a system with less than $2k$ independent equations and at least $2k$ variables. Hence, it has a nonzero solution, which gives the required $C \neq [0]$. ■

22.7 The Max-Flow Min-Cut Theorem

A *capacitated network* is a triple (V, E, k), where V is a finite set containing at least two distinguished elements $q, s \in V$ called *source* (q) and *sink* (s); E is a subset of $V \times V$ such that $v \neq w$, $v \neq s$, and $w \neq q$ for all $(v, w) \in E$; and $k : E \to \mathbb{R}_+$. Elements of V are called *vertices* and elements of E are called *edges*. The number

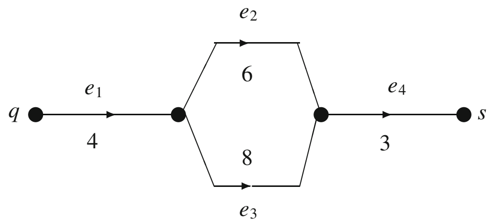

Fig. 22.2 A capacitated network

$k(e)$ is the *capacity* of the edge e; if $e = (v, w)$ then $k(e)$ is interpreted as the maximal amount that can flow from v to w through edge e. The source has only outgoing and the sink only incoming edges. See Fig. 22.2 for an example.

A *flow* in this network is a map $f : E \to \mathbb{R}_+$ with $f(e) \leq k(e)$ and such that for all $v \in V \setminus \{q, s\}$

$$\sum_{w \in V:\ (w,v) \in E} f(w, v) = \sum_{w \in V:\ (v,w) \in E} f(v, w) \ .$$

In other words, a flow satisfies the capacity constraints and for all vertices (except source and sink) the 'inflow' equals the 'outflow'.

The *value* of a flow f is defined as the inflow in the sink, i.e., as the number

$$\sum_{v \in V:\ (v,s) \in E} f(v, s) \ .$$

A flow is called *maximal* if it has maximal value among all possible flows. Intuitively, the value of a maximal flow is determined by the 'bottlenecks' in the network. In order to formalize this, define a *cut* in the network to be a partition of V into two subsets V_1 and V_2 such that $q \in V_1$ and $s \in V_2$. Such a cut is denoted by (V_1, V_2). The *capacity* of a cut is the number

$$k(V_1, V_2) := \sum_{v \in V_1, w \in V_2 : (v,w) \in E} k(v, w) \ ,$$

i.e., the total capacity along edges going from V_1 to V_2. A cut is called *minimal* if it has minimal capacity among all possible cuts.

In the example in Fig. 22.2, a minimal cut has only the sink in V_2, and its capacity is equal to 3. Obviously, this is also the value of a maximal flow, but such a flow is not unique.

Flows and cuts are, first of all, related as described in the following lemma.

Lemma 22.15 *Let f be a flow in the capacitated network (V, E, k), and let $\varphi : E \to \mathbb{R}$ be an arbitrary function. Then:*

(a) $\sum_{v \in V} \sum_{(w,v):\ (w,v) \in E} \varphi(w, v) = \sum_{v \in V} \sum_{(v,w):\ (v,w) \in E} \varphi(v, w) \ .$

(b) $\sum_{(q,v):\ (q,v)\in E} f(q,v) = \sum_{(v,s):\ (v,s)\in E} f(v,s)$.

(c) *For every cut* (V_1, V_2) *the value of the flow* f *is equal to*

$$\sum_{(v,w):\ (v,w)\in E, v\in V_1, w\in V_2} f(v,w) - \sum_{(v,w):\ (v,w)\in E, v\in V_2, w\in V_1} f(v,w) .$$

Proof (a) follows because summation at both sides is taken over the same sets. Part (a) moreover implies

$$\sum_{(q,v):\ (q,v)\in E} f(q,v) + \sum_{(v,w)\in E:\ v\neq q} f(v,w) =$$
$$\sum_{(v,s):\ (v,s)\in E} f(v,s) + \sum_{(v,w)\in E:\ w\neq s} f(v,w)$$

which implies (b) because $\sum_{(v,w)\in E:\ v\neq q} f(v,w) = \sum_{(v,w)\in E:\ w\neq s} f(v,w)$ by definition of a flow ('inflow' equals 'outflow' at every vertex that is not the source and not the sink). For part (c), let (V_1, V_2) be a cut of the network. Then

$$\begin{aligned}
\sum_{(v,w)\in E:\ v\in V_1, w\in V_2} f(v,w) &= \sum_{(v,w)\in E:\ v\in V_1} f(v,w) - \sum_{(v,w)\in E:\ v,w\in V_1} f(v,w) \\
&= \sum_{(v,w)\in E:\ v=q} f(v,w) + \sum_{(v,w)\in E:\ w\in V_1} f(v,w) \\
&\quad - \sum_{(v,w)\in E:\ v,w\in V_1} f(v,w) \\
&= \sum_{(v,w)\in E:\ v=q} f(v,w) + \sum_{(v,w)\in E:\ v\in V_2, w\in V_1} f(v,w) \\
&= \sum_{(v,w)\in E:\ w=s} f(v,w) + \sum_{(v,w)\in E:\ v\in V_2, w\in V_1} f(v,w) ,
\end{aligned}$$

where the second equality follows since everything that leaves from a node of V_1 also has entered that node, except for q; and the fourth equality follows from (b). This implies part (c) of the lemma. ■

The following theorem is the famous Max Flow Min Cut Theorem.

Theorem 22.16 *Let* (V, E, k) *be a capacitated network. Then the value of a maximal flow is equal to the capacity of a minimal cut.*

Proof Let f be a maximal flow. (Note that f is an optimal solution of a feasible bounded linear program, so that existence of f is guaranteed.) Part (c) of

Lemma 22.15 implies that the value of *any* flow f is smaller than or equal to

$$\sum_{(v,w):\ (v,w)\in E, v\in V_1, w\in V_2} k(v,w) - \sum_{(v,w):\ (v,w)\in E, v\in V_2, w\in V_1} f(v,w)$$
$$\leq \sum_{(v,w):\ (v,w)\in E, v\in V_1, w\in V_2} k(v,w)$$

for *any* cut (V_1, V_2), so that it is sufficient to find a cut of which the capacity is equal to the value of f.

For points v, w in the network define a *path* as a sequence of different non-directed edges starting in v and ending in w; 'non-directed' means that for any edge $(x, y) \in E$, (x, y) as well as (y, x) may be used in this path. Such a path may be described by a sequence $v = x_1, x_2, \ldots, x_m = w$ with $(x_i, x_{i+1}) \in E$ or $(x_{i+1}, x_i) \in E$ for every $i = 1, \ldots, m-1$. Call such a path *non-satiated* if for every $i = 1, \ldots, m-1$ it holds that $f(x_i, x_{i+1}) < k(x_i, x_{i+1})$ if $(x_i, x_{i+1}) \in E$, and $f(x_{i+1}, x_i) > 0$ if $(x_{i+1}, x_i) \in E$. In other words, the flow is below capacity in edges that are traversed in the 'right' way, and positive in edges that are traversed in the 'wrong' way.

Define V_1 to be the set of vertices x for which there is a non-satiated path from q to x, together with the vertex q, and let V_2 be the complement of V_1 in V. Then $s \in V_2$ because otherwise there would be a non-satiated path from q to s, implying that f would not be maximal; the flow f could be increased by increasing it in edges on this path that are traversed in the right way and decreasing it in edges along the path that are traversed in the wrong way, without violating the capacity constraints or the inflow-outflow equalities. Hence (V_1, V_2) is a cut in the network.

Let $(x, y) \in E$ with $x \in V_1$ and $y \in V_2$. Then $f(x, y) = k(x, y)$ because otherwise there would be a non-satiated path from q to a vertex in V_2. Similarly, $f(x', y') = 0$ whenever $(x', y') \in E$ with $x' \in V_2$ and $y' \in V_1$. By Lemma 22.15, part (c), the value of the flow f is equal to

$$\sum_{(v,w):\ (v,w)\in E, v\in V_1, w\in V_2} f(v,w) - \sum_{(v,w):\ (v,w)\in E, v\in V_2, w\in V_1} f(v,w)$$

hence to $\sum_{(v,w):\ (v,w)\in E, v\in V_1, w\in V_2} k(v,w)$, which is by definition the capacity of the cut (V_1, V_2). This completes the proof. ■

Observe that the proof of Theorem 22.16 suggests an algorithm to determine a maximal flow, by starting with an arbitrary flow, looking for a non-satiated path, and improving this path. By finding an appropriate cut, maximality of a flow can be checked. Theorem 22.16 is actually a (linear) duality result, but the above proof is elementary.

22.8 Problems

22.1. *Convex Sets*
Prove that a closed set $Z \subseteq \mathbb{R}^n$ is convex if and only if $\frac{1}{2}\mathbf{x} + \frac{1}{2}\mathbf{y} \in Z$ for all $\mathbf{x}, \mathbf{y} \in Z$.

22.2. *Proof of Lemma 22.3*
Prove that at most one of the systems in Lemma 22.3 has a solution.

22.3. *Rank of AA^T*
Let A be an $m \times n$ matrix with rank k. Prove that the rank of AA^T (where A^T is the transpose of A) is also equal to k.

22.4. *Proof of Lemma 22.5*
Prove that at most one of the systems in Lemma 22.5 has a solution.

22.5. *Proof of Lemma 22.7*
Prove Lemma 22.7.

22.6. *Extreme Points*
Let C be a convex set in a linear space V and let $e \in C$. Prove that the following three statements are equivalent.

(a) $e \in \text{ext}(C)$.
(b) For all $0 < \alpha < 1$ and all $x, y \in C$, if $x \neq y$ then $e \neq \alpha x + (1 - \alpha)y$.
(c) $C \setminus \{e\}$ is a convex set.

22.7. *Affine Subspaces*
Prove that a subset A of a linear space V is affine if, and only if, with any two different elements x and y of A, also the straight line through x and y is contained in A.

22.8. *The Set of Sup-points of a Linear Function on a Convex Set*
Let $f : \mathbb{R}^n \to \mathbb{R}^n$ be a linear function. Let C be a convex subset of $\mathbb{R}^n$ and $\alpha := \sup\{f(\mathbf{x}) \mid \mathbf{x} \in C\}$, $D := \{\mathbf{x} \in C \mid f(\mathbf{x}) = \alpha\}$. Show that D is convex and that $\text{ext}(D) \subseteq \text{ext}(C)$.

22.9 Notes

Theorem 22.10 is due to Brouwer (1912). For proofs, see also e.g. Scarf (1973) or Vohra (2005).

Theorem 22.11 is from Kakutani (1941). One way to prove this theorem is to derive it from the Brouwer Fixed Point Theorem: see, e.g., Hildenbrand and Kirman (1976).

A good source for results on convexity is Rockafellar (1970).

The Max Flow Min Cut Theorem (Theorem 22.16) is due to Ford and Fulkerson (1956).

References

Brouwer, L. E. J. (1912). Über Abbildung von Mannigfaltigkeiten. *Mathematische Annalen, 71*, 97–115.

Ford, L. R., & Fulkerson, D. R. (1956). Maximal flow through a network. *Canadian Journal of Mathematics, 8*, 399–404.

Hildenbrand, H., & Kirman, A. P. (1976). *Introduction to equilibrium analysis*. Amsterdam: North-Holland.

Kakutani, S. (1941). A generalization of Brouwer's fixed point theorem. *Duke Mathematical Journal, 8*, 457–459.

Rockafellar, R. T. (1970). *Convex analysis*. Princeton: Princeton University Press

Scarf, H. E. (1973). *The computation of economic equilibria*. New Haven: Cowles Foundation, Yale University Press.

Vohra, R. V. (2005). *Advanced mathematical economics*. New York: Routledge.