Finite Algorithms in Optimization and Data Analysis

M. R. Osborne
Department of Statistics,
Australian National University,
Canberra

Library of Congress Cataloging in Publication Data:

Osborne, M. R. (Michael Robert)
Finite algorithms in optimization and data analysis.

(Wiley series in probability and mathematical
statistics. Applied probability and statistics)
Includes bibliographies and index.
1. Mathematical optimization—Data processing.
2. Least squares—Data processing. 3. Approximation
theory—Data processing. 4. Algorithms. I. Title.
II. Series.
QA402.5.08 1984 519 84-11841

ISBN 0 471 90539 9

British Library Cataloguing in Publication Data:

Osborne, M. R.
Finite algorithms in optimization and data
analysis.—(Wiley series in probability and
mathematical statistics. Applied
probability and statistics)
1. Algorithms
I. Title
511'.8 QA9.58

ISBN 0 471 90539 9

Printed and bound in Great Britain.

Contents

CHAPTER 1

Some results from convex analysis

1.1 INTRODUCTION

This chapter summarizes primarily material relating to convex optimization problems in order to provide a suitable underpinning for subsequent algorithmic developments. The problem domain is one in which convexity is important (for example there is a sense in which linear programming is our archetypal problem), and which can be characterized formally as

$$\min_{\mathbf{x}\in S} f(\mathbf{x}) \tag{1.1}$$

where S is a prescribed convex set (the constraint set), and f is convex on $S \subseteq R^p$.

To motivate the kind of results to be considered, note that the structure of S is clearly of relevance, and this leads directly to consideration of representation theorems for convex sets. These results have a direct and elegant application to linear programming problems. The second class of results concerns separation theorems and their direct relation to the characterization of optimizing points. To see this, note that if the minimizing value in (1.1) is $f=\bar{f}$ then

$$S \cap \{\mathbf{x}; f(\mathbf{x}) < \bar{f}\} = \varnothing. \tag{1.2}$$

As both sets involved are convex there exists a separating hyperplane, and from the equation of this hyperplane we can deduce necessary conditions for an optimum (Kuhn–Tucker conditions). An interesting feature of this approach to characterizing optima is that it does not require f to be differentiable.

If $\mathbf{x} \in S$, $f(\mathbf{x}) > \bar{f}$ then there exists a direction $\mathbf{t}$ at $\mathbf{x}$ such that $\mathbf{x}+\gamma\mathbf{t} \in S$ provided $\gamma > 0$ is small enough, and such that f decreases in the direction $\mathbf{t}$. Such a direction is called a *direction of descent* for f at $\mathbf{x}$. Theorems of the alternative arise naturally in this context as either a downhill direction exists or $\mathbf{x}$ is an optimum. Convex functions need not be differentiable (an example is $|x|$ at $x=0$), but a generalized set-valued derivative called the subdifferential can be defined if the epigraph of f, epi f, the convex set lying above the graph of f in R^{p+1}, possesses a nonvertical supporting hyperplane at $\mathbf{x}$. The subdifferential is important in characterizing descent directions and in developing multiplier relations giving necessary conditions for $\mathbf{x}$ to be an optimum. The related concept of the directional derivative of a convex function will prove to be an important tool in analysing the structure of polyhedral convex functions. Also, we note that considering the supporting hyperplanes to epi f leads naturally to the conjugate convex function $f^*(\mathbf{u})$ and to the idea of duality which greatly enhances the structural richness of the problem setting. Finally, a brief description of descent methods for minimizing a convex function is given.

1.2 CONVEX SETS

In this section necessary material on the representation of convex sets and on separation theorems is developed. The key result on the existence of a hyperplane separating two disjoint convex sets is used extensively in subsequent sections. It has direct application to the development of alternative theorems and of necessary conditions for solutions of optimization problems. The representation theorem states that a convex set, under certain mild conditions, can be described completely in terms of quantities which have a natural geometric importance (extreme points and directions of recession). In the linear programming context this theorem has the direct interpretation that the optimum must be obtained at a vertex of the feasible region (an extreme point of this convex set), and that the problem can have a bounded solution only if the directions of recession bear a particular relationship to the objective function. Linear programming (discussed in Chapter 2) and the problems discussed in Chapters 3 and 4 are all particular cases of problems involving polyhedral convex functions. Polyhedral convexity occurs when there are only a finite number of extreme points and directions of recession, and the opportunity is taken here to develop some of the basic ideas.

Definition 2.1 The set $S \subseteq R^p$ is *convex* if

$$\mathbf{x}, \mathbf{y} \in S \Rightarrow \theta\mathbf{x} + (1-\theta)\mathbf{y} \in S \quad \text{for} \quad 0 \leq \theta \leq 1. \tag{2.1}$$

Equivalently, S is convex if all finite convex combinations of points in S is again in S. That is,

$$\mathbf{x}_i \in S, i = 1, 2, \ldots, m, \quad \sum_{i=1}^{m} \lambda_i = 1, \quad \lambda_i \geq 0, 1 \leq m < \infty \Rightarrow \sum_{i=1}^{m} \lambda_i \mathbf{x}_i \in S \tag{2.2}$$

Association with any set S is the set obtained by taking all convex combinations of points of S in the sense expressed by (2.2). This set is called the *convex hull* of S and is written conv S.

Example 2.1 Any function $\|\cdot\|$ taking bounded values for finite $\mathbf{x}$ is a *norm* if it satisfies the conditions:

(i) $\|\mathbf{x}\|>0, \quad \mathbf{x}\neq 0,$
(ii) $\|\mathbf{x}+\mathbf{y}\|\leq\|\mathbf{x}\|+\|\mathbf{y}\|,$ the triangle inequality, and
(iii) $\|\sigma\mathbf{x}\|=|\sigma|\,\|\mathbf{x}\|$.

Define $S=\{\mathbf{x};\|\mathbf{x}\|\leq 1\}$. Then it follows from (ii), (iii) that S is convex. Note that S is *balanced* ($\mathbf{x}\in S \Rightarrow -\mathbf{x}\in S$), and has a proper interior as $\theta\mathbf{x}/\|\mathbf{x}\|\in S$, $-1\leq\theta\leq 1$ for any $\mathbf{x}$. It is instructive to sketch S when $p=2$ for the particular cases

$$\|\mathbf{x}\|=\max|x_i|, \quad \left\{\sum_{i=1}^{2} x_i^2\right\}^{1/2}, \quad \sum_{i=1}^{2}|x_i|,$$

corresponding to the maximum, Euclidean, and l_1 norms respectively.

Remark 2.1 Alternatively, given S satisfying the above requirements, a norm can be defined by

$$\|\mathbf{x}\|_S=\inf\lambda, \qquad \mathbf{x}\in\lambda S. \tag{2.3}$$

If S is not balanced then the resulting function does not satisfy (iii) but is still convex. It is called a gauge function.

Definition 2.2 A *hyperplane* is the set of points

$$H(\mathbf{u},\nu)=\{\mathbf{x};\mathbf{u}^T\mathbf{x}=\nu\}. \tag{2.4}$$

It should be noted that $H(\mathbf{u},\nu)$ *separates* R^p into two distinct half-spaces

$$H^+(\mathbf{u},\nu)=\{\mathbf{x};\mathbf{u}^T\mathbf{x}>\nu\}, \tag{2.4a}$$

and

$$H^-(\mathbf{u},\nu)=\{\mathbf{x};\mathbf{u}^T\mathbf{x}\leq\nu\}. \tag{2.4b}$$

Lemma 2.1 (*Lemma of the separating hyperplane – simplest case*). *Let S be a closed convex set in R^p and $\mathbf{x}_0$ a point not in S. Then there exists a hyperplane H separating $\mathbf{x}_0$ and S in the sense that $S\subset H^+$, $\mathbf{x}_0\in H^-$.*

Proof Let $\mathbf{x}_1$ be any point in S. Then in the Euclidean vector norm

$$\inf_{\mathbf{x}\in S}\|\mathbf{x}-\mathbf{x}_0\|_2^2\leq\|\mathbf{x}_1-\mathbf{x}_0\|_2^2=r^2$$

The function $\|\mathbf{x}-\mathbf{x}_0\|_2^2$ is continuous on the closed set $S\cap\{\mathbf{x};\|\mathbf{x}-\mathbf{x}_0\|_2\leq r\}$ so the minimum value for $\mathbf{x}\in S$ is attained for $\mathbf{x}=\mathbf{x}^*$ (say). Let $\mathbf{y}\in S$ and consider $\mathbf{x}^*+\gamma\mathbf{z}$ where $\mathbf{z}=\mathbf{y}-\mathbf{x}^*$ and $\gamma>0$. Then

$$\|\mathbf{x}^*+\gamma\mathbf{z}-\mathbf{x}_0\|_2^2=\|\mathbf{x}^*-\mathbf{x}_0\|_2^2+2\gamma\mathbf{z}^T(\mathbf{x}^*-\mathbf{x}_0)+\gamma^2\|\mathbf{z}\|_2^2.$$

Letting $\gamma\to 0$ gives

$$\mathbf{z}^T(\mathbf{x}^*-\mathbf{x}_0)\geq 0, \qquad \mathbf{y}\in S. \tag{2.5}$$

It follows that $H(\mathbf{x}^*-\mathbf{x}_0, \mathbf{x}^{*T}(\mathbf{x}^*-\mathbf{x}_0)-\varepsilon)$ is a suitable hyperplane for all $\varepsilon>0$ small enough (ε must be chosen smaller than $\|\mathbf{x}_0-\mathbf{x}^*\|_2^2$).

Remark 2.2 Separation in this form is called *strong* separation, and there is even a small enough ball about $\mathbf{x}^*$ which excludes $\mathbf{x}_0$. A more subtle result is also important. It can be proved by induction in finite-dimensional spaces, but requires the Hahn–Banach theorem in more general situations.

Theorem 2.1 *Let S, T be convex sets, and $S\cap T=\varnothing$. Then there exists a hyperplane such that $S\subset H^+$, $T\subset H^-$. Equivalently we can find $\mathbf{u}$ such that*

$$\inf_{\mathbf{x}\in S} \mathbf{u}^T\mathbf{x} \geqslant \sup_{\mathbf{x}\in T} \mathbf{u}^T\mathbf{x} \tag{2.6}$$

Remark 2.3 An important application corresponds to S an *open* convex set, and $\mathbf{x}_0\in \text{cl}\, S\backslash S$. The theorem shows there is a hyperplane through $\mathbf{x}_0$ containing S in H^+. It follows that H contains only boundary points of S.

Definition 2.3 The hyperplane H in Remark 2.3 *supports* S at $\mathbf{x}_0$. Note that there exists a supporting hyperplane at every finite boundary point of S.

Definition 2.4 The function

$$\delta^*(\mathbf{u}\mid S)=\sup_{\mathbf{x}\in S} \mathbf{u}^T\mathbf{x} \tag{2.7}$$

is called the *support* function for S.

If the supremum in (2.7) is attained at $\mathbf{x}_0$, then $\mathbf{x}_0$ is a boundary point of S, $H(\mathbf{u}, \mathbf{u}^T\mathbf{x}_0)$ supports S at $\mathbf{x}_0$, and $S\subseteq H^-$. For compatibility with the support function, unless otherwise indicated, the convention will be followed that if H supports S then $S\subseteq H^-$.

Definition 2.5 The point $\mathbf{x}_0$ is an *extreme point* of S if and only if it cannot be expressed as a point properly in the interior of the line segment joining two distinct points of cl S. An extreme point which has the property that there exists a hyperplane supporting S at $\mathbf{x}_0$ such that

$$H\cap \text{cl}\, S=\{\mathbf{x}_0\} \tag{2.8}$$

is called an *exposed point.* Exposed points are equivalent to extreme points if the set of extreme points is finite.

Example 2.2 Consider $S=\{\mathbf{x}; \|\mathbf{x}\|\leqslant 1\}$.

(a) If $\|\mathbf{x}\|=\left\{\sum_{i=1}^{p} x_i^2\right\}^{1/2}$ then every point in cl $S\backslash$int S is an extreme point.

(b) If $\|\mathbf{x}\| = \max_{1 \le i \le p} |x_i|$ then the extreme points of S have the form

$$\mathbf{x} = \sum_{i=1}^{p} \theta_i \mathbf{e}_i, \qquad \theta_i = \pm 1, \; i = 1, 2, \ldots, p.$$

(c) If $\|\mathbf{x}\| = \sum_{i=1}^{p} |x_i|$ then the extreme points of S have the form

$$\mathbf{x} = \pm \mathbf{e}_1, \pm \mathbf{e}_2, \ldots, \pm \mathbf{e}_p.$$

In each case note that any $\mathbf{x} \in S$ can be written as a convex combination of extreme points. These are particular examples of a general result which will now be developed.

Lemma 2.2 *Let H support S at $\mathbf{x}$. If $\mathbf{y}$ is an extreme point of $H \cap S$ then $\mathbf{y}$ is an extreme point of S.*

Proof If $H \cap S = \{\mathbf{y}\}$ then $\mathbf{y}$ is an extreme point of S. Thus we assume that $H \cap S = T$, not a singleton, and that $\mathbf{y}$ is an extreme point of T but not of S. Then we can find $\mathbf{x}, \mathbf{z}$ in S but not in T such that

$$\mathbf{y} = \theta \mathbf{x} + (1 - \theta)\mathbf{z}, \qquad 0 < \theta < 1.$$

Now $\mathbf{y} \in H$, but $\mathbf{x}, \mathbf{z} \in \operatorname{int} H^-$. Thus

$$0 = \mathbf{u}^T\mathbf{y} - \nu = \theta(\mathbf{u}^T\mathbf{x} - \nu) + (1 - \theta)(\mathbf{u}^T\mathbf{z} - \nu) < 0$$

This gives a contradiction.

Lemma 2.3 *Let S be a closed, bounded, convex set. Then S has extreme points.*

Proof This is by induction with respect to dimension. If S is a singleton then the result is immediate. Now assume $S \subset R^p$, $\mathbf{x}_0$ is a boundary point of S, and H is a hyperplane supporting S at $\mathbf{x}_0$. It follows that $H \cap S$ is bounded, so that by the induction hypothesis it has extreme points. But then, by Lemma 2.2, these points are extreme points of S.

The representation theorem for bounded sets can now be given.

Theorem 2.2 *A closed bounded convex set S in R^p is the closed convex hull of its extreme points.*

Proof Let $\hat{S}$ be the closed convex hull of the extreme points of S. Then $\hat{S} \subseteq S$. Assume that $\hat{S} \subset S$ properly so that there exists a point $\mathbf{x} \in S$ strongly separated from $\hat{S}$. The separating hyperplane theorem now gives $H(\mathbf{u}, \nu)$

such that $\hat{S} \subset H^-$, $\mathbf{x} \in H^+$. Consider

$$\nu_0 = \delta^*(\mathbf{u} \mid S) > \nu$$

It follows from the definition of the support function that $H(\mathbf{u}, \nu_0)$ supports S. Also $T = H(\mathbf{u}, \nu_0) \cap S$ is closed and bounded. Thus, by Lemma 2.3, it has an extreme point. By Lemma 2.2 this is also an extreme point of S. But by construction it is strongly separated from $\hat{S}$, giving a contradiction.

To extend this result to unbounded sets a further concept is needed in order to describe the property that points can be arbitrarily far apart.

Definition 2.6 Let S be a convex set, and let $\mathbf{t}$ have the property that

$$\mathbf{x} \in S \Rightarrow \mathbf{x} + \lambda \mathbf{t} \in S, \qquad \lambda \geq 0 \tag{2.9}$$

then $\mathbf{t}$ is a *direction of recession* for S. It is an *extreme direction* if it cannot be represented as a convex combination of other directions (so that there do not exist directions $\mathbf{t}_i$ and constants $\lambda_i \geq 0$, $i = 1, 2, \ldots, k$ such that $\mathbf{t} = \sum_{i=1}^{k} \lambda_i \mathbf{t}_i$ for any finite k). S is said to *contain a line* if both $\mathbf{t}$ and $-\mathbf{t}$ are directions of recession.

It is not difficult to give examples of important convex sets containing directions of recession.

Definition 2.7 Let $\mathbf{b} \in R^m$, $A : R^p \to R^m$, then

$$M = \{\mathbf{x};\, A\mathbf{x} = \mathbf{b}\} \tag{2.10}$$

is an *affine set* or *flat* in R^p.

Thus every hyperplane is an affine set, and every affine set is an intersection of hyperplanes. An affine set can be represented as a translated subspace

$$M = \mathbf{x}_0 + L \tag{2.11}$$

where $\mathbf{x}_0$ is any point such that

$$A\mathbf{x}_0 = \mathbf{b} \tag{2.12}$$

and

$$L = \{\mathbf{y};\, A\mathbf{y} = 0\} \tag{2.13}$$

is a linear space. The dimension of M is the dimension of $L = p - \text{rank}\,(A)$.

Remark 2.4 A form of the separation theorem that will be required is that if M is an affine set, S convex, and $S \cap M = \varnothing$, then there exists a separating hyperplane containing M.

Example 2.3 (a) Any element of $L \neq 0$ defines a direction of recession for M. Any set of the form $S+L$ where S is convex and L a subspace has no extreme points. (b) A convex set $C \subseteq R^p$ is a *cone* pointed at $\mathbf{x}_0$ if

$$\mathbf{y} \in C \Rightarrow \mathbf{x}_0 + \lambda(\mathbf{y}-\mathbf{x}_0) \in C \qquad \forall\, \lambda \geq 0. \tag{2.14}$$

Thus $\mathbf{y}-\mathbf{x}_0$ is a direction of recession for $\mathbf{y}$ in C, $\mathbf{y} \neq \mathbf{x}_0$, and $\mathbf{x}_0$ is the only point in C that can be an extreme point. In particular, it is contained in every supporting hyperplane to C (either $\delta^*(\mathbf{u} \mid C) = \mathbf{u}^T\mathbf{x}_0$ if $\mathbf{u}^T(\mathbf{y}-\mathbf{x}_0) \leq 0$, $\forall\, \mathbf{y} \in C$, or $\delta^*(\mathbf{u} \mid C) = +\infty$). An important example is $R_p^+ = \{\mathbf{x};\, x_i \geq 0\}$ which is a cone pointed at 0.

We now state the extended representation theorem. The standard proofs of this result use induction explicitly with respect to the dimension (in the proof of Theorem 2.2 this was hidden in the appeal to Lemma 2.3). This extended result is particularly important in providing a theoretical basis for linear programming.

Theorem (*Klee*) *A closed convex set in R^p containing no lines is the convex hull of its extreme points and directions of recession. That is, given $\mathbf{x} \in S$, then there exist extreme points of S, $\mathbf{s}_1, \mathbf{s}_2, \ldots, \mathbf{s}_k$, and extreme directions of recession, $\mathbf{t}_1, \mathbf{t}_2, \ldots, \mathbf{t}_l$, such that*

$$\mathbf{x} = \sum_{i=1}^{k} \theta_i \mathbf{s}_i + \sum_{i=1}^{l} \lambda_i \mathbf{t}_i \tag{2.15}$$

where

$$\sum_{i=1}^{k} \theta_i = 1, \quad \theta_i \geq 0, i = 1, 2, \ldots, k, \quad \lambda_i \geq 0, i = 1, 2, \ldots, l.$$

Definition 2.8 If the set of extreme points and extreme directions is finite then S is *polyhedral*. If S is bounded then it is a *convex polyhedron* or a bounded *convex polytope*.

A convex polyhedron with a proper interior possesses a representation as the intersection of a finite number of closed half-spaces. These are determined by the subsets of $\geq p$ extreme points which determine a hyperplane such that the convex polyhedron is contained in one or the other of the closed half-spaces so generated.

A convex set is a *convex body* if it has an interior point. If it does not contain interior points then it is contained in some affine set.

Definition 2.9 The intersection of all affine sets containing S is called the *affine hull* of S and is denoted aff S. It is the smallest affine set containing S.

Definition 2.10 The *relative interior* of the convex set S (ri S) is

$$\text{ri}\, S = \{\mathbf{x};\, \mathbf{x} \in \text{aff}\, S, \exists \varepsilon > 0 \ni \varepsilon B(\mathbf{x}) \cap \text{aff}\, S \subset S\} \tag{2.16}$$

where the Kronecker delta δ_{ij} is defined by

$$\delta_{ij} = \begin{cases} 1 \ (i = j) \\ 0 \ (i \neq j) \end{cases} \tag{1.2.11}$$

The set of vectors $\{\mathbf{x}_i \in R^n \mid i = 1, \ldots, m; m \leqslant n\}$ is a *linearly independent* set if and only if

$$\sum_{i=1}^{m} c_i \mathbf{x}_i = \mathbf{0} \tag{1.2.12}$$

implies that $c_i = 0$ $(i = 1, \ldots, m)$, and is *linearly dependent* otherwise. A set of vectors $\{\mathbf{u}_i \in R^n \mid i = 1, \ldots, m; m \geqslant n\}$ *spans* or *generates* the linear space R^n if an only if every vector $\mathbf{x} \in R^n$ is expressible in the form

$$\mathbf{x} = \sum_{i=1}^{m} c_i \mathbf{u}_i \tag{1.2.13}$$

A *basis* for the linear space R^n is a linearly independent set of vectors in R^n which spans R^n. The set of vectors $\{\mathbf{e}_i \in R^n \mid i = 1, \ldots, n\}$ where

$$\mathbf{e}_i = [0, \ldots, 0, \overset{(i)}{1}, 0, \ldots, 0]^T \quad (i = 1, \ldots, n) \tag{1.2.14}$$

in which $\mathbf{e}_i$ has all components equal to zero save the ith which is equal to unity, clearly forms a basis for R^n since if $\mathbf{x} = [x_1, \ldots, x_n]^T$ then

$$\mathbf{x} = \sum_{i=1}^{n} x_i \mathbf{e}_i \tag{1.2.15}$$

It can be shown that every basis for R^n contains exactly n vectors, and that every set of n orthogonal vectors in R^n excluding $\mathbf{0}$ forms a basis for R^n. An *orthonormal* basis for R^n is a set of n orthonormal vectors. For example, $\{\mathbf{e}_1, \ldots, \mathbf{e}_n\}$ defined by (1.2.14) is an orthonormal basis for R^n.

Given any basis for R^n, an orthonormal basis for R^n may be constructed by using the Gram–Schmidt procedure. Let $\{\mathbf{u}_1, \ldots, \mathbf{u}_n\}$ be a basis for R^n and let $\mathbf{v}_i$, $\mathbf{w}_i$ $(i = 1, \ldots, n)$ be generated recursively from

$$\mathbf{v}_1 = \mathbf{u}_1 \qquad \mathbf{w}_1 = \frac{\mathbf{v}_1}{(\mathbf{v}_1^T \mathbf{v}_1)^{1/2}} \tag{1.2.16a}$$

$$\mathbf{v}_i = \mathbf{u}_i - \sum_{k=1}^{i-1} (\mathbf{w}_k^T \mathbf{u}_k) \mathbf{w}_k \qquad \mathbf{w}_i = \frac{\mathbf{v}_i}{(\mathbf{v}_i^T \mathbf{v}_i)^{1/2}} \quad (i = 2, \ldots, n) \tag{1.2.16b}$$

Then it is easy to show (Exercise 1.2.7) that

$$\mathbf{w}_i^T \mathbf{w}_j = \delta_{ij} \quad (i, j = 1, \ldots, n) \tag{1.2.17}$$

so that $\{\mathbf{w}_1, \ldots, \mathbf{w}_n\}$ is an orthonormal basis for R^n.

A set $S \subset R^n$ is a *subspace* of the linear space R^n if and only if for each pair $\mathbf{x}'$, $\mathbf{x}''$ of vectors in S, the vector $\alpha\mathbf{x}' + \beta\mathbf{x}''$ is also in S where α, β are any real numbers. As with R^n, the dimensionality m of a subspace S of R^n is the number of vectors in any basis for S. Clearly $m \leqslant n$. Let $\{\mathbf{u}_i \in R^n \mid i = 1, \ldots, m\}$ be a basis for the m-dimensional subspace S of R^n. Then clearly an orthonormal basis for S can be constructed by using the Gram–Schmidt procedure (1.2.16), where i runs from 2 to m.

Exercise 2.1

(i) A closed convex set is the intersection of the closed half-spaces containing it.

(ii) Let C_1, C_2 be convex sets, then

$$\operatorname{cl} C_1 \subseteq \operatorname{cl} C_2 \quad \text{if and only if} \quad \delta^*(\mathbf{u} \mid C_1) \leq \delta^*(\mathbf{u} \mid C_2).$$

(iii) Let S be the set determined by the linear inequalities

$$x + y \geq 1,$$
$$10y - x \geq -1$$
$$-y + 10x \geq -1$$

Determine its extreme points and directions of recession. Also determine the hyperplanes supporting S at each extreme point.

1.3 THEOREMS OF THE ALTERNATIVE

Alternative theorems are important in function minimization as they permit the formalization of the alternatives that either the current point is optimal or there exists a direction in which the function can be reduced.

Separation theorems are the main analytic device used in this section. This is well illustrated by the first result which is perhaps the simplest form of alternative theorem and gives a criterion for the consistency of a system of linear inequalities.

Theorem 3.1 *Let S be a closed convex set. Then either the system of inequalities*

$$\mathbf{u}^T\mathbf{t} < 0, \qquad \mathbf{t} \in S, \tag{3.1}$$

is consistent or

$$0 \in S.$$

Proof The inequality (3.1) states that $\delta^*(\mathbf{u} \mid S) < 0$ so that 0 is strongly separated from S.

Remark 3.1 This result has meaning when S is not convex. Either (3.1) holds or $0 \in \operatorname{conv} S$.

Perhaps the most celebrated theorem of the alternative is known as Farkas' lemma, and it plays a key role in developing multiplier conditions characterizing optimality in mathematical programming problems. Here the separation theorem is applied to the particular case in which S is the cone generated by convex combinations of the rows of a matrix A. The resulting cone is polyhedral as the set of extreme directions is clearly no larger than the row dimension of A. Such a cone is also called *finitely generated*. A key preliminary result is that a finitely generated cone is closed.

Lemma 3.1 *Let the cone C be finitely generated by* $\mathbf{a}_1, \mathbf{a}_2, \ldots, \mathbf{a}_m$. *That is,*

$$C = \left\{\mathbf{x};\ \exists \lambda_i \geq 0,\ i = 1, 2, \ldots, m \ni \mathbf{x} = \sum_{i=1}^{m} \lambda_i \mathbf{a}_i\right\} \tag{3.2}$$

Then C is closed.

Proof C is convex by definition. However, the closure result is complicated if C contains lines. To exclude this case let L be the linear space generated by the lines in C, and introduce coordinates into L, $L^\perp$ such that

$$C = L \times L^\perp \cap C \tag{3.3}$$

As L is closed the result entails showing that the cone $L^\perp \cap C$ (which by construction does not contain lines) is closed. Thus there is no restriction in considering only this case ($L^\perp \cap C$ is clearly finitely generated).

If C does not contain lines then

$$\sum_{i=1}^{m} \lambda_i \mathbf{a}_i = 0 \Rightarrow \lambda_i = 0, \qquad i = 1, 2, \ldots, m$$

for if this is not the case then $\lambda_k > 0$ for some k so that both $\mathbf{a}_k$ and

$$-\mathbf{a}_k = \sum_{i \neq k} \frac{\lambda_i}{\lambda_k} \mathbf{a}_i$$

are in C, showing that C contains a line. It follows that $\mathbf{x}$ is bounded if and only if the λ_i in every representation of $\mathbf{x}$ are bounded, and closure is an immediate consequence.

Theorem 3.2 (*Farkas' lemma*) *Let* $A : R^p \rightarrow R^n$, $\mathbf{a}$ *be such that*

$$A\mathbf{x} \geq 0 \Rightarrow \mathbf{a}^T\mathbf{x} \geq 0. \tag{3.4}$$

Then there exists $\mathbf{y} \geq 0 \in R^n$ ($\mathbf{y} \in R_n^+$) *such that*

$$\mathbf{a}^T = \mathbf{y}^T A. \tag{3.5}$$

Proof This proceeds by constructing the cone

$$C = \left\{\mathbf{z};\ \mathbf{z} = \sum_{i=1}^{n} \lambda_i \mathbf{a}_i,\ \lambda_i \geq 0,\ \mathbf{a}_i = \boldsymbol{\rho}_i(A),\ i = 1, 2, \ldots, n\right\} \tag{3.6}$$

and assumes that $\mathbf{a} \notin C$. The separating hyperplane theorem is then used to give a contradiction, for by Lemma 3.1 C is closed so that $\mathbf{a}$ must be strongly separated from C. Choose the separating hyperplane to support C

so that it passes through $\mathbf{z}=0$. Then

$$C \subset H^{-}(\mathbf{u}, 0), \qquad \mathbf{a} \in H^{+}$$

But then choosing $\mathbf{x}=-\mathbf{u}$ in (3.4) gives a contradiction.

Remark 3.2 The idea of the alternative may be clearer if the result is put slightly differently. It says that exactly one of the pairs of systems

(i) $$A^{\mathrm{T}}\mathbf{y}=\mathbf{b}, \qquad \mathbf{y} \geqslant 0,$$

and

(i) $$A\mathbf{x} \geqslant 0, \qquad \mathbf{b}^{\mathrm{T}}\mathbf{x}<0,$$

can have a solution.

An alternative route to this result makes use of the *polar* to C. This is the cone

$$C^{*}=\{\mathbf{y} ; \mathbf{y}^{\mathrm{T}}\mathbf{x} \leqslant 0, \forall\, \mathbf{x} \in C\} \tag{3.7}$$

The chief properties of the polar cone are summarized below.

(a) C^{*} is a closed convex cone.
(b) If $C_1 \subseteq C_2$ then $C_2^{*} \subseteq C_1^{*}$.
(c) $C^{**}=C$ if and only if C is a closed convex cone.
(d) $C^{*}=(\mathrm{cl\,conv}\, C)^{*}$, the polar cone of the closure of the convex hull of C.
(e) If A is a linear space then $A^{\perp}=A^{*}$.
(f) If C_1 and C_2 are convex cones then

$$C_1^{*} \cap C_2^{*}=(C_1+C_2)^{*} \tag{3.8}$$

Farkas lemma follows from (3.8) for if C_1 is given by (3.6), and C_2 is generated by $\mathbf{a}$, then (3.4) gives

$$C_1^{*} \subseteq C_2^{*}$$

so that

$$C_1^{*}=C_1^{*} \cap C_2^{*}=(C_1+C_2)^{*}$$

and the result follows by taking polars and using (c).

Theorem 3.2 can be generalized substantially to partial orders defined on functions taking values in cones ($a \geqslant b$ if $a-b \in C$), to operators on normed linear spaces, and to convex (not only linear) operators. The following result is typical.

Theorem 3.3 *Let C be a convex cone in R^n, $S \subset R^p$ convex, and $A: R^p \rightarrow R^n$. Then exactly one of the systems*

(i) $$A\mathbf{x} \in \operatorname{int} C, \qquad \mathbf{x} \in S, \tag{3.9}$$

and

(ii) $$(\mathbf{v}^{\mathrm{T}}A)S \subset R_1^+, \qquad 0 \neq \mathbf{v} \in C^* \tag{3.10}$$

has a solution.

Proof It is clear that (i) and (ii) cannot both have a solution. If (i) has no solution then AS and int C are disjoint. But then the separating hyperplane theorem guarantees an $H(\mathbf{u}, 0)$ such that $AS \subset H^-$, int $C \subset H^+$. But this implies that

$$(\mathbf{u}^{\mathrm{T}}A)S \leqslant 0, \ \mathbf{u}^{\mathrm{T}}\mathbf{w} > 0, \ \forall\, \mathbf{w} \in \operatorname{int} C$$

so that $\mathbf{u} \in C^*$ and $\mathbf{u}$ solves (ii). This completes the proof of the theorem as either (i) has a solution or not, and if it does not then (ii) has a solution by the above argument.

Exercise 3.1

(i) Prove the properties (a)–(f) for polar cones.

(ii) Restate Farkas' lemma for cone-valued operators.

(iii) Prove Motzkin's theorem: Let $A: R^p \to R^n$, $B: R^p \to R^m$, $T \subset R^n$ a closed convex cone, $S \subset R^m$ a convex cone with int $S \neq \varnothing$. Then exactly one of the following systems has a solution

(a) $$-A\mathbf{x} \in T, \qquad -B\mathbf{x} \in \operatorname{int} S \tag{3.11}$$

(b) $$\mathbf{v}^{\mathrm{T}}B + \mathbf{w}^{\mathrm{T}}A = 0, \qquad \mathbf{w} \in T^*, \qquad 0 \neq \mathbf{v} \in S^*. \tag{3.12}$$

(iv) Deduce Farkas' lemma as a special case of Motzkin's theorem.

(v) Show that either

$\exists\, \mathbf{x} \ni A\mathbf{x} = \mathbf{c}$,

or

$\exists\, \mathbf{v} \ni A^{\mathrm{T}}\mathbf{v} = 0, \qquad \mathbf{v}^{\mathrm{T}}\mathbf{c} = 1.$

Exercise 3.2 Let $A: R^p \to R^n$ and $X = \{\mathbf{x}; A\mathbf{x} \geqslant \mathbf{b}\}$. The constraint $\boldsymbol{\rho}_i(A)\mathbf{x} \geqslant b_i$ is *redundant* if X is unchanged by deletion of this inequality.

(i) What is the maximum possible value of $\dim(X \cap \{\boldsymbol{\rho}_i(A)\mathbf{x} = b_i\})$.

(ii) Use Farkas lemma to determine conditions for a constraint to be redundant.

Distinguish between the cases that the constraint hyperplane contains or does not contain points of the feasible region X.

1.4 CONVEX FUNCTIONS

All the problems considered in Chapters 2 to 5 can be reduced to that of minimizing particular convex functions (frequently even polyhedral convex functions). Convexity is a strong assumption (for example, points at which a