26:711:564 **OPTIMIZATION MODELS IN FINANCE**
Professor Andrzej Ruszczyński

Convex Sets and Functions

1 Convex Sets

1.1 Basic Properties

The notion of a convex set is central to optimization theory. A convex set is such that, for any two of its points, the entire segment joining these points is contained in the set.

Definition 1.1 A set $X \subset \mathbb{R}^n$ is called *convex* if for all $x^1 \in X$ and $x^2 \in X$ it contains all points

$$\alpha x^1 + (1-\alpha)x^2, \quad 0 < \alpha < 1.$$

Convexity is preserved by the operation of intersection.

Lemma 1.2 *Let I be an arbitrary index set. If the sets $X_i \subset \mathbb{R}^n$, $i \in I$, are convex, then the set $X = \bigcap_{i \in I} X_i$ is convex.*

intersection of convex sets is convex set.

Sets can be subjected to algebraic operations similarly to points in $\mathbb{R}^n$. We can multiply a set X by a scalar c to get

$$cX = \{y \in \mathbb{R}^n : y = cx,\ x \in X\}.$$

The *Minkowski sum* of two sets is defined as follows:

$$X + Y = \{z \in \mathbb{R}^n : z = x + y,\ x \in X,\ y \in Y\}.$$

These operations preserve convexity.

Lemma 1.3 *Let X and Y be convex sets in $\mathbb{R}^n$ and let c and d be real numbers. Then the set $Z = cX + dY$ is convex.*

A point $\alpha x^1 + (1-\alpha)x^2$, where $\alpha \in [0,1]$, appearing in Definition 1.1, belongs to the segment joining x^1 and x^2. We can "join" more points by constructing their convex hull.

V.S. linear combination

Definition 1.4 A point x is called a *convex combination* of points $x^1, \ldots, x^m$ if there exist $\alpha_1 \geq 0, \ldots, \alpha_m \geq 0$ such that

$$x = \alpha_1 x^1 + \alpha_2 x^2 + \cdots + \alpha_m x^m$$

and

$$\alpha_1 + \alpha_2 + \cdots + \alpha_m = 1.$$

Definition 1.5 The *convex hull* of the set X (denoted by conv$X$) is the intersection of all convex sets containing X. smallest convex set of X

The relation between these two concepts is the subject of the next lemma.

Lemma 1.6 *The set* conv$X$ *is the set of all convex combinations of points of* X*.*

For example, the convex hull of three points in $\mathbb{R}^2$ is the triangle having these points as vertices.

In the above result we considered all convex combinations, that is, for all numbers m of points and for arbitrary selections of these points. Both m and the points selected can be restricted. We now show that the number of points m need not be larger than $n+1$. In Section 2 we further show that only some special points need to be considered, if the set X is compact.

Lemma 1.7 *If* $X \subset \mathbb{R}^n$*, then every element of* conv$X$ *is a convex combination of at most* $n+1$ *points of* X*.*

The last result is known as *Carathéodory's Theorem.*

2 Extreme Points

Definition 2.1 A point x of a convex set X is called an *extreme point* of X if no other points $x^1 \in X$ and $x^2 \in X$ exist such that

$$x = \frac{1}{2}x^1 + \frac{1}{2}x^2.$$

Extreme points of a compact convex set fully characterize the set.[1]

Theorem 2.2 *A convex and compact set in* $\mathbb{R}^n$ *is equal to the convex hull of the set of its extreme points.*

Our results are important for linear programming.

[1] Theorem 2.2 is known as *Minkowski's Theorem.*

Theorem 2.3 *Let A be an $m \times n$ matrix and let the set $X \subset \mathbb{R}^n$ be defined as*

$$X = \{x \in \mathbb{R}^n : Ax = b,\ x \geq 0\}.$$

A point x is an extreme point of X if and only if the columns of A that correspond to positive components of x are linearly independent.

Solutions of the system $Ax = b$ whose nonzero components correspond to linearly independent columns of A, are called in linear programming *basic solutions*. If they are nonnegative, they are called *basic feasible solutions*. Their role is now evident.

Theorem 2.4 *If the set $X = \{x \in \mathbb{R}^n : Ax = b,\ x \geq 0\}$ is bounded, it is the convex hull of the set of basic feasible solutions.*

3 Cones

3.1 Basic Concepts

A particular class of convex sets, convex cones, play a significant role in optimization theory.

Definition 3.1 A set $K \subset \mathbb{R}^n$ is called a *cone* if for every $x \in K$ and all $\alpha > 0$ one has $\alpha x \in K$. A *convex cone* is a cone that is a convex set.

A simple example of a convex cone in $\mathbb{R}^n$ is the nonnegative orthant:

$$\mathbb{R}^n_+ = \{x \in \mathbb{R}^n : x_j \geq 0,\ j = 1, \ldots, n\}.$$

For convex cones, positive combinations remain in the set, similar to convex combinations for convex sets.

Lemma 3.2 *Let K be a convex cone. If $x^1 \in K$, $x^2 \in K$, ..., $x^m \in K$ and $\alpha_1 > 0$, $\alpha_2 > 0, \ldots, \alpha_m > 0$, then $\alpha_1 x^1 + \alpha_2 x^2 + \cdots + \alpha_m x^m \in K$.*

Let us consider two important examples of convex cones.

Lemma 3.3 *Assume that X is a convex set. Then the set*

$$\operatorname{cone}(X) = \{\gamma x : x \in X,\ \gamma \geq 0\}$$

is a convex cone.

The set cone(X) is called the *cone generated by the set* X. For a convex set X and a point $x \in X$ the set

$$K_X(x) = \text{cone}(X - x)$$

is called the *cone of feasible directions* of X at x (or the *radial* cone). It follows directly from the definition that it is a convex cone.

Definition 3.4 Let K be a cone in $\mathbb{R}^n$. The set

$$K^\circ = \{y \in \mathbb{R}^n : \langle y, x \rangle \le 0 \text{ for all } x \in K\}$$

is called the *polar cone* of K.

For example, the set $K = \{x \in \mathbb{R}^n : x_i \ge 0, \ i = 1, \dots, n\}$ is a convex closed cone. Its polar is $-K$, as can be verified directly. The negative of the polar cone, $K^* = -K^\circ$, is called the *dual cone.*

We use the notation $K^{\circ\circ}$ for the polar cone of the polar cone of K. The following result is known as the *Bipolar Theorem.*

Theorem 3.5 *If $K \subset \mathbb{R}^n$ is a closed convex cone, then*

$$K^{\circ\circ} = K.$$

Polar cones to pre-images of cones under linear transformations can be calculated in an explicit form.

Theorem 3.6 *Assume that A is an $m \times n$ matrix. Let*

$$K = \{x \in \mathbb{R}^n : Ax \le 0\}.$$

Then K is a closed convex cone and

$$K^\circ = \{A^T \lambda : \lambda \ge 0\}.$$

This result is known as *Farkas Lemma*). It is frequently formulated as an alternative: *exactly one of the following two systems has a solution, either*

(i) $Ax \le 0$ and $\langle c, x \rangle > 0$; *or*

(ii) $c = A^T \lambda$, $\lambda \ge 0$.

Indeed, (i) is not solvable if and only if $c \in K^\circ$, which is equivalent to (ii).

Example 3.7 Suppose we have n securities with present prices $c_1, \dots, c_n$. The securities can be either bought or sold *short*, that is, borrowed and sold for cash. Any amounts can be traded.

At some future time one of m states may occur. The price of security j in state i will be equal to a_{ij}, $i = 1, \dots, m$, $j = 1, \dots, n$. At this time we will liquidate our holding, that is, the securities held will be sold, and the short positions will be covered by purchasing and returning the amounts borrowed.

An *arbitrage*, in the simplest version, is the existence of $\bar{x} \in \mathbb{R}^n$ such that $\langle c, \bar{x} \rangle < 0$ and $A\bar{x} \geq 0$. Such a portfolio $\bar{x}$ can be "purchased" with an immediate profit and then liquidated at a future point in time with no additional loss, whatever the future state. If arbitrage is present, considering portfolios $M\bar{x}$, with $M \to \infty$, we can increase the profits without any limits.

Defining the convex cone

$$K = \{x \in \mathbb{R}^n : Ax \geq 0\},$$

we see that the absence of arbitrage is equivalent to the fact that $\langle c, x \rangle \geq 0$ for all $x \in K$. The latter is nothing else but

$$-c \in K^\circ.$$

Farkas Lemma implies that there exist *state prices* $p \geq 0$ such that

$$c = A^T p. \tag{1}$$

The converse is obvious: if such state prices exist, then

$$\langle c, x \rangle = \langle A^T p, x \rangle = \langle p, Ax \rangle \geq 0 \quad \text{for all} \quad x \in \mathbb{R}^n.$$

Suppose the first security (cash) has a constant unit price of 1 for all states: $c_1 = 1$ and $a_{i1} = 1$, $i = 1, \dots, m$. Then it follows from the first relation in (1) that

$$\sum_{i=1}^{m} p_i = 1.$$

The vector p may be interpreted as a vector of implied probabilities of states $1, \dots, m$.

4 Convex Functions

4.1 Basic Concepts

In our analysis it is convenient to consider functions which may take, in addition to real values, two special values: $-\infty$ and $+\infty$. The real line augmented with these special values will be denoted by $\overline{\mathbb{R}}$.

With every function $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ we can associate two sets: the *domain*

$$\text{dom} f = \{x : f(x) < +\infty\}$$

and the *epigraph*

$$\text{epi}\, f = \{(x, v) \in \mathbb{R}^n \times \mathbb{R} : v \geq f(x)\}.$$

Definition 4.1 A function f is called *convex* if $\operatorname{epi} f$ is a convex set.

Definition 4.2 A function f is called *concave* if $-f$ is convex.

Lemma 4.3 *A function f is convex if and only if for all x^1 and x^2 and for all $0 \le \alpha \le 1$ we have*

$$f(\alpha x^1 + (1-\alpha)x^2) \le \alpha f(x^1) + (1-\alpha)f(x^2). \tag{2}$$

Inequality (2) can be used as an alternative definition of proper convex functions.

Definition 4.4 A function f is called *strictly convex* if inequality (2) is strict for all $x^1 \ne x^2$ and for all $0 < \alpha < 1$.

Lemma 4.5 *If f_i, $i \in I$, is a family of convex functions, then*

$$f(x) = \sup_{i \in I} f_i(x)$$

is convex.

Lemma 4.6 *If f is a convex function, then for all $x^1, x^2, \dots, x^m$ and all $\alpha_1 \ge 0, \alpha_2 \ge 0, \dots, \alpha_m \ge 0$ such that $\alpha_1 + \alpha_2 + \cdots + \alpha_m = 1$, one has*

$$f(\alpha_1 x^1 + \alpha_2 x^2 + \cdots + \alpha_m x^m) \le \alpha_1 f(x^1) + \alpha_2 f(x^2) + \cdots + \alpha_m f(x^m).$$

Lemma 4.7 *If the functions f_i, $i = 1, 2, \dots, m$, are convex, then for all $c_1 \ge 0, c_2 \ge 0, \dots, c_m \ge 0$ the function*

$$f(x) = c_1 f_1(x) + c_2 f_2(x) + \cdots + c_m f_m(x)$$

is convex.

Lemma 4.8 *If $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ is convex, then for each $\beta \in \mathbb{R}$ the set*

$$M_\beta = \{x : f(x) \le \beta\} \tag{3}$$

is convex. If, in addition, f is lower semicontinuous, then the set M_β is closed for all β.

The set M_β in the last lemma is called the *level set* of f.

Not every function that has convex level sets is convex, as the example of $f(x) = \sqrt{|x|}$ clearly demonstrates.

Lemma 4.9 *Let $X \subset \mathbb{R}^n$ be a convex set and let $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ be a convex function. Then the set $\hat{X}$ of solutions of the optimization problem*

$$\underset{x \in X}{\text{minimize}}\, f(x) \tag{4}$$

is convex.

The maxima of convex functions in convex sets can be characterized as well.

Theorem 4.10 *Let $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ be a convex function and let $X \subset \text{dom} f$ be a convex, closed and bounded set. Then the set of solutions of the problem*

$$\underset{x \in X}{\text{maximize}}\;\; f(x) \tag{5}$$

contains at least one extreme point of X. If, in addition, the function $f(\cdot)$ is affine, then the set of solutions of (5) *is the convex hull of the set of the extreme points of X that are solutions of* (5)*.*

We can now return to the linear programming problem, which we analyzed at the end of Section 2.

Theorem 4.11 *If the feasible set of the linear programming problem*

$$\begin{aligned} \text{minimize} \quad & \langle c, x \rangle \\ \textit{subject to} \quad & Ax = b, \\ & x \geq 0, \end{aligned} \tag{6}$$

is bounded, then the set of optimal solutions is the convex hull of the set of optimal basic feasible solutions.

The above fact is used by the simplex method for solving linear programming problems. It moves from one basic feasible solution to a better one, as long as progress is possible. The best basic feasible solution is guaranteed to be optimal. It can be found after finitely many steps if a solution exists. If the set is unbounded, we may discover a ray from the recession cone, along which the objective can be decreased without limits. In this case no optimal solution exists.

4.2 Smooth Convex Functions

We now formulate convexity criteria for smooth functions. We denote by $\nabla f(x)$ the *gradient* of the function f at x,

$$\nabla f(x) = \begin{bmatrix} \frac{\partial f(x)}{\partial x_1} \\ \frac{\partial f(x)}{\partial x_2} \\ \vdots \\ \frac{\partial f(x)}{\partial x_n} \end{bmatrix}.$$

Here $x_1, x_2, \ldots, x_n$ denote the coordinates of the vector x.

If f is twice continuously differentiable, $\nabla^2 f(x)$ denotes the *Hessian* of f at x,

$$\nabla^2 f(x) = \begin{bmatrix} \frac{\partial^2 f(x)}{\partial x_1^2} & \frac{\partial^2 f(x)}{\partial x_1 \partial x_2} & \cdots & \frac{\partial^2 f(x)}{\partial x_1 \partial x_n} \\ \frac{\partial^2 f(x)}{\partial x_2 \partial x_1} & \frac{\partial^2 f(x)}{\partial x_2^2} & \cdots & \frac{\partial^2 f(x)}{\partial x_2 \partial x_n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial^2 f(x)}{\partial x_n \partial x_1} & \frac{\partial^2 f(x)}{\partial x_n \partial x_2} & \cdots & \frac{\partial^2 f(x)}{\partial x_n^2} \end{bmatrix}.$$

We use the symbol $\mathcal{C}^2$ to denote the set of twice continuously differentiable real valued functions. For $f \in \mathcal{C}^2$ the Hessian is a symmetric matrix, because the values of the mixed derivatives do not depend on the order of differentiation.

Theorem 4.12 *Assume that a function f is continuously differentiable. Then*

(i) *f is convex if and only if for all x and y*

$$f(y) \geq f(x) + \langle \nabla f(x), y - x \rangle; \tag{7}$$

(ii) *f is strictly convex if and only if for all $x \neq y$*

$$f(y) > f(x) + \langle \nabla f(x), y - x \rangle. \tag{8}$$

Theorem 4.13 *Assume that $f : \mathbb{R}^n \to \mathbb{R}$ is twice continuously differentiable. Then*

(i) *f is convex if and only if its Hessian $\nabla^2 f(x)$ is positive semidefinite for all $x \in \mathbb{R}^n$;* 半正定矩阵

(ii) *if the Hessian $\nabla^2 f(x)$ is positive definite for all $x \in \mathbb{R}^n$ then f is strictly convex.* 正定矩阵：左上角的所有行列式都为正

Note that statement (ii) of the theorem does not have the "only if" part, which appeared in the quadratic case. For example, the function $f(x) = x^4$ is strictly convex, but its second derivative vanishes at 0.