1

Basic Convexity Concepts

Contents

In this chapter and the following three, we develop the theory of convex sets, which is the mathematical foundation for minimax theory, Lagrange multiplier theory, and duality. We assume no prior knowledge of the subject, and we give a detailed development. As we embark on the study of convexity, it is worth listing some of the properties of convex sets and functions that make them so special in optimization.

(a) *A convex function has no local minima that are not global.* Thus the difficulties associated with multiple disconnected local minima, whose global optimality is hard to verify in practice, are avoided (see Section 2.1).

(b) *A convex set has a nonempty relative interior.* In other words, relative to the smallest affine set containing it, a convex set has a nonempty interior (see Section 1.4). Thus convex sets avoid the analytical and computational optimization difficulties associated with "thin" and "curved" constraint surfaces.

(c) *A convex set is connected and has feasible directions at any point* (assuming it consists of more than one point). By this we mean that given any point x in a convex set X, it is possible to move from x along some directions y and stay within X for at least a nontrivial interval, i.e., $x + \alpha y \in X$ for all sufficiently small but positive stepsizes α (see Section 4.6). In fact a stronger property holds: given any two distinct points x and $\overline{x}$ in X, the direction $\overline{x} - x$ is a feasible direction at x, and all feasible directions can be characterized this way. For optimization purposes, this is important because it allows a calculus-based comparison of the cost of x with the cost of its close neighbors, and forms the basis for some important algorithms. Furthermore, much of the difficulty commonly associated with discrete constraint sets (arising for example in combinatorial optimization), is not encountered under convexity.

(d) *A nonconvex function can be "convexified" while maintaining the optimality of its global minima*, by forming the convex hull of the epigraph of the function (see Exercise 1.20).

(e) *The existence of a global minimum of a convex function over a convex set is conveniently characterized in terms of directions of recession* (see Section 2.3).

(f) *A polyhedral convex set (one that is specified by linear equality and inequality constraints) is characterized in terms of a finite set of extreme points and extreme directions.* This is the basis for finitely terminating methods for linear programming, including the celebrated simplex method (see Sections 3.3 and 3.4).

(g) *A convex function is continuous within the interior of its domain, and has nice differentiability properties.* In particular, a real-valued

convex function is directionally differentiable at any point. Furthermore, while a convex function need not be differentiable, it possesses subgradients, which are nice and geometrically intuitive substitutes for a gradient (see Chapter 4). Just like gradients, subgradients figure prominently in optimality conditions and computational algorithms.

(h) *Convex functions are central in duality theory.* Indeed, the dual problem of a given optimization problem (discussed in Chapter 6) consists of minimization of a convex function over a convex set, even if the original problem is not convex.

(i) *Closed convex cones are self-dual with respect to polarity.* In words, we have $C = (C^*)^*$ for any closed and convex cone C, where C^* is the polar cone of C (the set of vectors that form a nonpositive inner product with all vectors in C), and $(C^*)^*$ is the polar cone of C^*. This simple and geometrically intuitive property (discussed in Section 3.1) underlies important aspects of Lagrange multiplier theory.

(j) *Convex lower semicontinuous functions are self-dual with respect to conjugacy.* It will be seen in Chapter 7 that a certain geometrically motivated conjugacy operation on a convex, lower semicontinuous function generates another convex, lower semicontinuous function, and when applied for the second time regenerates the original function. The conjugacy operation relies on a fundamental dual characterization of a closed convex set: as the union of the closures of all line segments connecting its points, and as the intersection of the closed halfspaces within which the set is contained. Conjugacy is central in duality theory, and has a nice interpretation that can be used to visualize and understand some of the most interesting aspects of convex optimization.

In this first chapter, after an introductory first section, we focus on the basic concepts of convex analysis: characterizations of convex sets and functions, convex and affine hulls, topological concepts such as closure, continuity, and relative interior, and the important notion of the recession cone.

1.1 LINEAR ALGEBRA AND REAL ANALYSIS

In this section, we list some basic definitions, notational conventions, and results from linear algebra and real analysis. We assume that the reader is familiar with this material, so no proofs are given. For related and additional material, we recommend the books by Hoffman and Kunze [HoK71], Lancaster and Tismenetsky [LaT85], and Strang [Str76] (linear algebra),

and the books by Ash [Ash72], Ortega and Rheinboldt [OrR70], and Rudin [Rud76] (real analysis).

Set Notation

If X is a set and x is an element of X, we write $x \in X$. A set can be specified in the form $X = \{x \mid x \text{ satisfies } P\}$, as the set of all elements satisfying property P. The union of two sets X_1 and X_2 is denoted by $X_1 \cup X_2$ and their intersection by $X_1 \cap X_2$. The symbols $\exists$ and $\forall$ have the meanings "there exists" and "for all," respectively. The empty set is denoted by $\emptyset$.

The set of real numbers (also referred to as scalars) is denoted by $\Re$. The set $\Re$ augmented with $+\infty$ and $-\infty$ is called the *set of extended real numbers*. We write $-\infty < x < \infty$ for all real numbers x, and $-\infty \leq x \leq \infty$ for all extended real numbers x. We denote by $[a, b]$ the set of (possibly extended) real numbers x satisfying $a \leq x \leq b$. A rounded, instead of square, bracket denotes strict inequality in the definition. Thus $(a, b]$, $[a, b)$, and (a, b) denote the set of all x satisfying $a < x \leq b$, $a \leq x < b$, and $a < x < b$, respectively. Furthermore, we use the natural extensions of the rules of arithmetic: $x \cdot 0 = 0$ for every extended real number x, $x \cdot \infty = \infty$ if $x > 0$, $x \cdot \infty = -\infty$ if $x < 0$, and $x + \infty = \infty$ and $x - \infty = -\infty$ for every scalar x. The expression $\infty - \infty$ is meaningless and is never allowed to occur.

Inf and Sup Notation

The *supremum* of a nonempty set X of scalars, denoted by $\sup X$, is defined as the smallest scalar y such that $y \geq x$ for all $x \in X$. If no such scalar exists, we say that the supremum of X is ∞. Similarly, the *infimum* of X, denoted by $\inf X$, is defined as the largest scalar y such that $y \leq x$ for all $x \in X$, and is equal to $-\infty$ if no such scalar exists. For the empty set, we use the convention

$$\sup \emptyset = -\infty, \qquad \inf \emptyset = \infty.$$

If $\sup X$ is equal to a scalar $\overline{x}$ that belongs to the set X, we say that $\overline{x}$ is the *maximum point* of X and we write $\overline{x} = \max X$. Similarly, if $\inf X$ is equal to a scalar $\overline{x}$ that belongs to the set X, we say that $\overline{x}$ is the *minimum point* of X and we write $\overline{x} = \min X$. Thus, when we write $\max X$ (or $\min X$) in place of $\sup X$ (or $\inf X$, respectively), we do so just for emphasis: we indicate that it is either evident, or it is known through earlier analysis, or it is about to be shown that the maximum (or minimum, respectively) of the set X is attained at one of its points.

Function Notation

If f is a function, we use the notation $f : X \mapsto Y$ to indicate the fact that f is defined on a nonempty set X (its *domain*) and takes values in a set Y (its *range*). Thus when using the notation $f : X \mapsto Y$, we implicitly assume that X is nonempty. If $f : X \mapsto Y$ is a function, and U and V are subsets of X and Y, respectively, the set $\{f(x) \mid x \in U\}$ is called the *image* or *forward image of* U *under* f, and the set $\{x \in X \mid f(x) \in V\}$ is called the *inverse image of* V *under* f.

1.1.1 Vectors and Matrices

We denote by $\Re^n$ the set of n-dimensional real vectors. For any $x \in \Re^n$, we use x_i to indicate its ith *coordinate*, also called its ith *component*.

Vectors in $\Re^n$ will be viewed as column vectors, unless the contrary is explicitly stated. For any $x \in \Re^n$, x' denotes the transpose of x, which is an n-dimensional row vector. The *inner product* of two vectors $x, y \in \Re^n$ is defined by $x'y = \sum_{i=1}^n x_i y_i$. Two vectors $x, y \in \Re^n$ satisfying $x'y = 0$ are called *orthogonal*.

If x is a vector in $\Re^n$, the notations $x > 0$ and $x \geq 0$ indicate that all components of x are positive and nonnegative, respectively. For any two vectors x and y, the notation $x > y$ means that $x - y > 0$. The notations $x \geq y$, $x < y$, etc., are to be interpreted accordingly.

If X is a set and λ is a scalar, we denote by λX the set $\{\lambda x \mid x \in X\}$. If X_1 and X_2 are two subsets of $\Re^n$, we denote by $X_1 + X_2$ the set

$$\{x_1 + x_2 \mid x_1 \in X_1,\, x_2 \in X_2\},$$

which is referred to as the *vector sum of* X_1 *and* X_2. We use a similar notation for the sum of any finite number of subsets. In the case where one of the subsets consists of a single vector $\overline{x}$, we simplify this notation as follows:

$$\overline{x} + X = \{\overline{x} + x \mid x \in X\}.$$

We also denote by $X_1 - X_2$ the set

$$\{x_1 - x_2 \mid x_1 \in X_1,\, x_2 \in X_2\}.$$

Given sets $X_i \subset \Re^{n_i}$, $i = 1, \ldots, m$, the *Cartesian product* of the X_i, denoted by $X_1 \times \cdots \times X_m$, is the set

$$\{(x_1, \ldots, x_m) \mid x_i \in X_i,\, i = 1, \ldots, m\},$$

which is a subset of $\Re^{n_1 + \cdots + n_m}$.

Subspaces and Linear Independence

A nonempty subset S of $\Re^n$ is called a *subspace* if $ax + by \in S$ for every $x, y \in S$ and every $a, b \in \Re$. An *affine set* in $\Re^n$ is a translated subspace, i.e., a set X of the form $X = \overline{x} + S = \{\overline{x} + x \mid x \in S\}$, where $\overline{x}$ is a vector in $\Re^n$ and S is a subspace of $\Re^n$, called the *subspace parallel to* X. Note that there can be only one subspace S associated with an affine set in this manner. [To see this, let $X = x + S$ and $X = \overline{x} + \overline{S}$ be two representations of the affine set X. Then, we must have $x = \overline{x} + \overline{s}$ for some $\overline{s} \in \overline{S}$ (since $x \in X$), so that $X = \overline{x} + \overline{s} + S$. Since we also have $X = \overline{x} + \overline{S}$, it follows that $S = \overline{S} - \overline{s} = \overline{S}$.] The *span* of a finite collection $\{x_1, \ldots, x_m\}$ of elements of $\Re^n$ is the subspace consisting of all vectors y of the form $y = \sum_{k=1}^m \alpha_k x_k$, where each α_k is a scalar.

The vectors $x_1, \ldots, x_m \in \Re^n$ are called *linearly independent* if there exists no set of scalars $\alpha_1, \ldots, \alpha_m$, at least one of which is nonzero, such that $\sum_{k=1}^m \alpha_k x_k = 0$. An equivalent definition is that $x_1 \neq 0$, and for every $k > 1$, the vector x_k does not belong to the span of $x_1, \ldots, x_{k-1}$.

If S is a subspace of $\Re^n$ containing at least one nonzero vector, a *basis* for S is a collection of vectors that are linearly independent and whose span is equal to S. Every basis of a given subspace has the same number of vectors. This number is called the *dimension* of S. By convention, the subspace $\{0\}$ is said to have dimension zero. The *dimension of an affine set* $\overline{x} + S$ is the dimension of the corresponding subspace S. Every subspace of nonzero dimension has a basis that is orthogonal (i.e., any pair of distinct vectors from the basis is orthogonal).

Given any set X, the set of vectors that are orthogonal to all elements of X is a subspace denoted by $X^\perp$:

$$X^\perp = \{y \mid y'x = 0, \forall\, x \in X\}.$$

If S is a subspace, $S^\perp$ is called the *orthogonal complement* of S. Any vector x can be uniquely decomposed as the sum of a vector from S and a vector from $S^\perp$. Furthermore, we have $(S^\perp)^\perp = S$.

Matrices

For any matrix A, we use A_{ij}, $[A]_{ij}$, or a_{ij} to denote its ijth element. The *transpose* of A, denoted by A', is defined by $[A']_{ij} = a_{ji}$. For any two matrices A and B of compatible dimensions, the transpose of the product matrix AB satisfies $(AB)' = B'A'$.

If X is a subset of $\Re^n$ and A is an $m \times n$ matrix, then the *image of* X *under* A is denoted by AX (or $A \cdot X$ if this enhances notational clarity):

$$AX = \{Ax \mid x \in X\}.$$

If Y is a subset of $\Re^m$, the *inverse image of* Y *under* A is denoted by $A^{-1}Y$ or $A^{-1} \cdot Y$:

$$A^{-1}Y = \{x \mid Ax \in Y\}.$$

If X and Y are subspaces, then AX and $A^{-1}Y$ are also subspaces.

Let A be a square matrix. We say that A is *symmetric* if $A' = A$. We say that A is *diagonal* if $[A]_{ij} = 0$ whenever $i \neq j$. We use I to denote the identity matrix (the diagonal matrix whose diagonal elements are equal to 1). We denote the *determinant* of A by $\det(A)$.

Let A be an $m \times n$ matrix. The *range space* of A, denoted by $R(A)$, is the set of all vectors $y \in \Re^m$ such that $y = Ax$ for some $x \in \Re^n$. The *nullspace* of A, denoted by $N(A)$, is the set of all vectors $x \in \Re^n$ such that $Ax = 0$. It is seen that the range space and the null space of A are subspaces. The *rank* of A is the dimension of the range space of A. The rank of A is equal to the maximal number of linearly independent columns of A, and is also equal to the maximal number of linearly independent rows of A. The matrix A and its transpose A' have the same rank. We say that A has *full rank*, if its rank is equal to $\min\{m, n\}$. This is true if and only if either all the rows of A are linearly independent, or all the columns of A are linearly independent.

The range space of an $m \times n$ matrix A is equal to the orthogonal complement of the nullspace of its transpose, i.e.,

$$R(A) = N(A')^{\perp}.$$

Another way to state this result is that given vectors $a_1, \ldots, a_n \in \Re^m$ (the columns of A) and a vector $x \in \Re^m$, we have $x'y = 0$ for all y such that $a_i'y = 0$ for all i if and only if $x = \lambda_1 a_1 + \cdots + \lambda_n a_n$ for some scalars $\lambda_1, \ldots, \lambda_n$. This is a special case of Farkas' Lemma, an important result for constrained optimization, which will be discussed in Section 3.2. A useful application of this result is that if S_1 and S_2 are two subspaces of $\Re^n$, then

$$S_1^{\perp} + S_2^{\perp} = (S_1 \cap S_2)^{\perp}.$$

This follows by introducing matrices B_1 and B_2 such that $S_1 = \{x \mid B_1 x = 0\} = N(B_1)$ and $S_2 = \{x \mid B_2 x = 0\} = N(B_2)$, and writing

$$S_1^{\perp} + S_2^{\perp} = R\left(\begin{bmatrix} B_1' & B_2' \end{bmatrix}\right) = N\left(\begin{bmatrix} B_1 \\ B_2 \end{bmatrix}\right)^{\perp} = \big(N(B_1) \cap N(B_2)\big)^{\perp} = (S_1 \cap S_2)^{\perp}$$

A function $f : \Re^n \mapsto \Re$ is said to be *affine* if it has the form $f(x) = a'x + b$ for some $a \in \Re^n$ and $b \in \Re$. Similarly, a function $f : \Re^n \mapsto \Re^m$ is said to be *affine* if it has the form $f(x) = Ax + b$ for some $m \times n$ matrix A and some $b \in \Re^m$. If $b = 0$, f is said to be a *linear function* or *linear transformation*. Sometimes, with slight abuse of terminology, an equation or inequality involving a linear function, such as $a'x = b$ or $a'x \leq b$, is referred to as a *linear equation or inequality*, respectively.

1.1.2 Topological Properties

Definition 1.1.1: A *norm* $\|\cdot\|$ on $\Re^n$ is a function that assigns a scalar $\|x\|$ to every $x \in \Re^n$ and that has the following properties:

(a) $\|x\| \geq 0$ for all $x \in \Re^n$.

(b) $\|\alpha x\| = |\alpha| \cdot \|x\|$ for every scalar α and every $x \in \Re^n$.

(c) $\|x\| = 0$ if and only if $x = 0$.

(d) $\|x + y\| \leq \|x\| + \|y\|$ for all $x, y \in \Re^n$ (this is referred to as the *triangle inequality*).

The *Euclidean norm* of a vector $x = (x_1, \ldots, x_n)$ is defined by

$$\|x\| = (x'x)^{1/2} = \left(\sum_{i=1}^{n} |x_i|^2 \right)^{1/2}.$$

We will use the Euclidean norm almost exclusively in this book. In particular, *in the absence of a clear indication to the contrary,* $\|\cdot\|$ *will denote the Euclidean norm.* Two important results for the Euclidean norm are:

Proposition 1.1.1: (Pythagorean Theorem) For any two vectors x and y that are orthogonal, we have

$$\|x + y\|^2 = \|x\|^2 + \|y\|^2.$$

Proposition 1.1.1: (Pythagorean Theorem) For any two vectors x and y that are orthogonal, we have

$$\|x + y\|^2 = \|x\|^2 + \|y\|^2.$$

Two other important norms are the *maximum norm* $\|\cdot\|_\infty$ (also called *sup-norm* or ℓ_∞*-norm*), defined by

$$\|x\|_\infty = \max_{i=1,\ldots,n} |x_i|,$$

and the ℓ_1-*norm* $\|\cdot\|_1$, defined by

$$\|x\|_1 = \sum_{i=1}^{n} |x_i|.$$

Sequences

We use both subscripts and superscripts in sequence notation. Generally, we prefer subscripts, but we use superscripts whenever we need to reserve the subscript notation for indexing components of vectors and functions. The meaning of the subscripts and superscripts should be clear from the context in which they are used.

A sequence $\{x_k \mid k = 1, 2, \ldots\}$ (or $\{x_k\}$ for short) of scalars is said to *converge* if there exists a scalar x such that for every $\epsilon > 0$ we have $|x_k - x| < \epsilon$ for every k greater than some integer K (that depends on ϵ). The scalar x is said to be the *limit* of $\{x_k\}$, and the sequence $\{x_k\}$ is said to *converge to* x; symbolically, $x_k \to x$ or $\lim_{k\to\infty} x_k = x$. If for every scalar b there exists some K (that depends on b) such that $x_k \geq b$ for all $k \geq K$, we write $x_k \to \infty$ and $\lim_{k\to\infty} x_k = \infty$. Similarly, if for every scalar b there exists some integer K such that $x_k \leq b$ for all $k \geq K$, we write $x_k \to -\infty$ and $\lim_{k\to\infty} x_k = -\infty$. Note, however, that implicit in any of the statements "$\{x_k\}$ converges" or "the limit of $\{x_k\}$ exists" or "$\{x_k\}$ has a limit" is that the limit of $\{x_k\}$ is a scalar.

A scalar sequence $\{x_k\}$ is said to be *bounded above* (respectively, *below*) if there exists some scalar b such that $x_k \leq b$ (respectively, $x_k \geq b$) for all k. It is said to be *bounded* if it is bounded above and bounded below. The sequence $\{x_k\}$ is said to be monotonically *nonincreasing* (respectively, *nondecreasing*) if $x_{k+1} \leq x_k$ (respectively, $x_{k+1} \geq x_k$) for all k. If $x_k \to x$ and $\{x_k\}$ is monotonically nonincreasing (nondecreasing), we also use the notation $x_k \downarrow x$ ($x_k \uparrow x$, respectively).

Proposition 1.1.3: Every bounded and monotonically nonincreasing or nondecreasing scalar sequence converges.

Note that a monotonically nondecreasing sequence $\{x_k\}$ is either bounded, in which case it converges to some scalar x by the above proposition, or else it is unbounded, in which case $x_k \to \infty$. Similarly, a monotonically nonincreasing sequence $\{x_k\}$ is either bounded and converges, or it is unbounded, in which case $x_k \to -\infty$.

Given a scalar sequence $\{x_k\}$, let

$$y_m = \sup\{x_k \mid k \geq m\}, \qquad z_m = \inf\{x_k \mid k \geq m\}.$$

The sequences $\{y_m\}$ and $\{z_m\}$ are nonincreasing and nondecreasing, respectively, and therefore have a limit whenever $\{x_k\}$ is bounded above or is bounded below, respectively (Prop. 1.1.3). The limit of y_m is denoted by $\limsup_{k\to\infty} x_k$, and is referred to as the *upper limit* of $\{x_k\}$. The limit of z_m is denoted by $\liminf_{k\to\infty} x_k$, and is referred to as the *lower limit* of $\{x_k\}$. If $\{x_k\}$ is unbounded above, we write $\limsup_{k\to\infty} x_k = \infty$, and if it is unbounded below, we write $\liminf_{k\to\infty} x_k = -\infty$.

Proposition 1.1.4: Let $\{x_k\}$ and $\{y_k\}$ be scalar sequences.

(a) We have

$$\inf\{x_k \mid k \geq 0\} \leq \liminf_{k\to\infty} x_k \leq \limsup_{k\to\infty} x_k \leq \sup\{x_k \mid k \geq 0\}.$$

(b) $\{x_k\}$ converges if and only if

$$-\infty < \liminf_{k\to\infty} x_k = \limsup_{k\to\infty} x_k < \infty.$$

Furthermore, if $\{x_k\}$ converges, its limit is equal to the common scalar value of $\liminf_{k\to\infty} x_k$ and $\limsup_{k\to\infty} x_k$.

(c) If $x_k \leq y_k$ for all k, then

$$\liminf_{k\to\infty} x_k \leq \liminf_{k\to\infty} y_k, \qquad \limsup_{k\to\infty} x_k \leq \limsup_{k\to\infty} y_k.$$

(d) We have

$$\liminf_{k\to\infty} x_k + \liminf_{k\to\infty} y_k \leq \liminf_{k\to\infty}(x_k + y_k),$$

$$\limsup_{k\to\infty} x_k + \limsup_{k\to\infty} y_k \geq \limsup_{k\to\infty}(x_k + y_k).$$

A sequence $\{x_k\}$ of vectors in $\Re^n$ is said to converge to some $x \in \Re^n$ if the ith component of x_k converges to the ith component of x for every i. We use the notations $x_k \to x$ and $\lim_{k\to\infty} x_k = x$ to indicate convergence for vector sequences as well. The sequence $\{x_k\}$ is called bounded if each of its corresponding component sequences is bounded. It can be seen that $\{x_k\}$ is bounded if and only if there exists a scalar c such that $\|x_k\| \leq c$ for all k. An infinite subset of a sequence $\{x_k\}$ is called a *subsequence* of $\{x_k\}$. Thus a subsequence can itself be viewed as a sequence, and can be represented as a set $\{x_k \mid k \in \mathcal{K}\}$, where $\mathcal{K}$ is an infinite subset of positive