Elements of Convex Optimization Theory

Costis Skiadas*

August 2015

This is a revised and extended version of Appendix A of Skiadas (2009), providing a self-contained overview of elements of convex optimization theory. The presentation is limited to the finite-dimensional case, but from an advanced perspective that is amenable to generalization. Some results, like the projection theorem and the Riesz representation theorem for Hilbert spaces, are stated and proved in a way that also applies to infinite-dimensional spaces, since doing so comes at negligible additional cost. On the other hand, instances where the finite dimensionality assumption is essential are noted.

1 Vector Space

A (real) **vector** or **linear space** is a set X, whose elements are called **vectors** or **points**, together with two operations, **addition** and **multiplication by scalars**, that satisfy the following conditions, for all vectors x, y, z and real numbers α, β :

1. To (x, y) addition assigns a unique element of X denoted by $x + y$, and to (α, x) multiplication by scalars assigns a unique element of X denoted by αx.
2. $x + y = y + x$, $x + (y + z) = (x + y) + z$ and $\alpha(\beta x) = (\alpha\beta)x$.
3. $\alpha(x + y) = \alpha x + \alpha y$ and $(\alpha + \beta)x = \alpha x + \beta x$.
4. $0 + x = x$ and $1x = x$, where 0 is an element of X called the **zero vector.**
5. There exists a vector $-x$ such that $x + (-x) = 0$.

*Kellogg School of Management, Department of Finance, Northwestern University, 2001 Sheridan Road, Evanston, IL 60208.

It follows that the zero vector is unique, and for each x in X, the vector $-x$ is unique. We write $x - y = x + (-y)$.

An underlying vector space X is taken as given throughout this appendix. Although the term "vector space X" is common, it should be emphasized that a vector space specification includes not only a set of vectors but also the rules for adding vectors and multiplying by scalars.

Example 1 *$X = \mathbb{R}^d$ is the usual d-dimensional Euclidean space. Vector addition and multiplication by scalars are defined coordinatewise: $(\alpha x + y)_i = \alpha x_i + y_i$ for every scalar α, vectors x, y and coordinate i.*

Example 2 *Given an underlying probability space, the set of all random variables of finite variance is a vector space, with the usual rules for adding and scaling random variables state by state.*

A subset L of X is a **linear** (or **vector**) **subspace** if it is a vector space itself or, equivalently, if for all $x, y \in L$ and $\alpha \in \mathbb{R}$, $x + y$ and αx are elements of L. The linear subspace **generated** or **spanned** by a set of vectors S, denoted by $\operatorname{span}(S)$, is the intersection of all linear subspaces that include S. Alternatively, the span of a set can be constructed from within: For a finite set of vectors $S = \{x_1, \ldots, x_n\}$, the set $\operatorname{span}(S)$, also denoted by $\operatorname{span}(x_1, \ldots, x_n)$, consists of all linear combinations of the form $\alpha_1 x_1 + \cdots + \alpha_n x_n$, where $\alpha_1, \ldots, \alpha_n \in \mathbb{R}$. For every set of vectors S, $\operatorname{span}(S) = \bigcup \{\operatorname{span}(F) : F \text{ is a finite subset of } S\}$.

A set of vectors S is **linearly independent** if every $x \in \operatorname{span}(S)$ has a unique representation of the form $x = \alpha_1 x_1 + \cdots + \alpha_n x_n$, for some $\alpha_1, \ldots, \alpha_n \in \mathbb{R}$ and $x_1, \ldots, x_n \in S$. It follows easily that the set of vectors S is linearly independent if and only if for all $x_1, \ldots, x_n \in S$ and $\alpha_1, \ldots, \alpha_n \in \mathbb{R}$,

$$\alpha_1 x_1 + \cdots + \alpha_n x_n = 0 \quad \text{implies} \quad \alpha_1 = \cdots = \alpha_n = 0.$$

A **basis** of X is a linearly independent set of vectors S that generates X, that is, $\operatorname{span}(S) = X$.

A vector space is **finite dimensional** if it has a finite basis and **infinite dimensional** otherwise. Example 2 with an infinite state space motivates our general interest in infinite-dimensional vector spaces. In this text, every state space is assumed to be finite, and therefore random variables can be encoded as finite-dimensional Euclidean vectors. We therefore focus on the finite-dimensional case, but as noted in the introductory remarks, from a perspective that allows direct extensions to contexts such as that of Example 2 with an infinite state space. Every basis of a finite-dimensional vector space has the same number of elements, called the space's **dimension.** The vector space $\{0\}$ has, by definition, dimension zero.

If X has finite dimension d, we represent a basis $\{B_1, \ldots, B_d\}$ of X as a column matrix

$$B = (B_1, \ldots, B_d)',$$

and we write σ^x to denote the row vector in $\mathbb{R}^d$ that **represents** the point x in X relative to the given basis B, meaning that

$$x = \sigma^x B = \sum_{i=1}^{d} \sigma_i^x B_i.$$

A **functional** is a function of the form $f : X \to \mathbb{R}$. A functional f is **linear** if $f(\alpha x + \beta y) = \alpha f(x) + \beta f(y)$ for all $x, y \in X$ and $\alpha, \beta \in \mathbb{R}$. Given a finite basis $B = (B_1, \ldots, B_d)'$ and a linear functional f, we use the notation

$$f(B) = (f(B_1), \ldots, f(B_d))'$$

for the column matrix that lists the values that f assigns to the basis elements. The single vector $f(B)$ in $\mathbb{R}^d$ determines the entire function f, since $x = \sigma^x B$ implies $f(x) = \sigma^x f(B)$.

A subset C of X is **convex** if for all $x, y \in C$, $\alpha \in (0,1)$ implies $\alpha x + (1-\alpha) y \in C$. The function $f : D \to \mathbb{R}$, where $D \subseteq X$, is **convex** if D is a convex set and for all $x, y \in D$, $\alpha \in (0,1)$ implies $f(\alpha x + (1-\alpha) y) \leq \alpha f(x) + (1-\alpha) f(y)$. The function f is **concave** if $-f$ is convex. A subset C of X is a **cone** if for all $x \in C$, $\alpha \in \mathbb{R}_+$ implies $\alpha x \in C$. One can easily check that a cone C is convex if and only if $x, y \in C$ implies $x + y \in C$, and that a convex cone C is a linear subspace if and only if $x \in C$ implies $-x \in C$.

An important type of convex set is a linear manifold. Given any subset S of X and vector x, we write $x + S = S + x = \{x + s : s \in S\}$ to denote the **translation** of S by x. A subset M of X is a **linear manifold** if some translation of M is a linear subspace. The **dimension** of a linear manifold M is that of the linear subspace to which M translates. An exercise shows:

Proposition 3 *The set $M \subseteq X$ is a linear manifold if and only if $x, y \in M$ and $\alpha \in \mathbb{R}$ implies $\alpha x + (1-\alpha) y \in M$.*

2 Inner Product

A widely applicable form of optimization can be thought of as a projection operation. Informally, the projection of a point x onto a convex set S is the point of S that is closest to x (assuming such a point exists). The simplest nontrivial instance of this problem is the case in which S is a line. Suppose that $S = \text{span}(y)$ and y has unit "length." The scalar s such that sy is the projection of x onto S is the inner product of x and y, denoted by $(y \mid x)$. Note that the functional $(y \mid \cdot)$ is linear. Defining $(\alpha y \mid x) = \alpha\,(y \mid x)$ for all $\alpha \in \mathbb{R}$, it is not hard to see that the function $(\cdot \mid \cdot)$ is bilinear (linear in each argument), symmetric ($(x \mid y) = (y \mid x)$) and positive definite ($(x \mid x) > 0$ for $x \neq 0$). With this geometric motivation, we take the notion of an inner product as a primitive object satisfying certain axioms, and (in later sections) we use inner products to characterize projections on convex sets more generally.

Definition 4 *A (real)* ***inner product*** $(\cdot \mid \cdot)$ *on the vector space X is a mapping that assigns to each $(x, y) \in X \times X$ a real number, denoted by $(x \mid y)$, and satisfies, for all $x, y, z \in X$ and $\alpha \in \mathbb{R}$,*

1. $(x + y \mid z) = (x \mid z) + (y \mid z)$ *and* $(\alpha x \mid y) = \alpha\,(x \mid y)$.
2. $(x \mid y) = (y \mid x)$.
3. $(x \mid x) \geq 0$, *with equality holding if and only if* $x = 0$.

An ***inner product space*** *is a vector space together with an inner product on this space.*

Example 5 *Suppose $X = \mathbb{R}^d$ and Q is a positive definite symmetric matrix. Arranging the elements of X as row vectors, the quadratic form $(x \mid y) = xQy'$ defines an inner product on X. If Q is the identity matrix, we obtain the* ***Euclidean inner product****: $(x \mid y) = x \cdot y = \sum_{i=1}^{d} x_i y_i$.*

Example 6 *Suppose X is any finite-dimensional vector space and B is any basis of X. Then an inner product is defined by letting $(x \mid y) = \sigma^x \cdot \sigma^y$, where $x = \sigma^x B$ and $y = \sigma^y B$. We will later establish the interesting fact that given any inner product on X, we can select the basis B so that this example's representation is valid.*

Example 7 *Suppose X is the set of all random variables on some finite probability space in which every state is assigned a positive probability mass. Then $(x \mid y) = \mathbb{E}\,[xy]$ defines* *an inner product. If Y is a linear subspace of X that does not contain the constant random variables, then $(x \mid y) = \operatorname{cov}[x, y] = \mathbb{E}[xy] - \mathbb{E}x\mathbb{E}y$ defines an inner product on Y but not on X (why?).*

We henceforth take as given the inner product space $(X, (\cdot \mid \cdot))$, which in the remainder of this section is assumed to be finite dimensional.

It will be convenient to extend the inner product notation to matrices of vectors, using the usual matrix addition and multiplication rules. In particular, given a column matrix of vectors $B = (B_1, \ldots, B_d)'$ and any vector x, we write

$$\left(x \mid B'\right) = \left((x \mid B_1), \ldots, (x \mid B_d)\right),$$

and

$$\left(B \mid B'\right) = \begin{pmatrix} (B_1 \mid B_1) & (B_1 \mid B_2) & \cdots & (B_1 \mid B_d) \\ (B_2 \mid B_1) & (B_2 \mid B_2) & \cdots & (B_2 \mid B_d) \\ \vdots & \vdots & \ddots & \vdots \\ (B_d \mid B_1) & (B_d \mid B_2) & \cdots & (B_d \mid B_d) \end{pmatrix}.$$

The matrix $(B \mid B')$ is known as the **Gram matrix** of B and plays a crucial role in the computation of projections. The following proposition shows that Gram matrices can be used to convert abstract inner-product notation to concrete expressions involving only matrices of scalars.

Proposition 8 *Suppose $B = (B_1, \ldots, B_d)'$ is a basis of X.*

(a) For any $x = \sigma^x B$ and $y = \sigma^y B$,

$$(x \mid y) = \sigma^x \left(B \mid B'\right) \sigma^{y\prime}.$$

(b) The Gram matrix $(B \mid B')$ is symmetric and positive definite.

(c) For any $x \in X$, the representation $x = \sigma^x B$ can be computed as

$$\sigma^x = \left(x \mid B'\right) \left(B \mid B'\right)^{-1}.$$

Proof. (a) Using the bilinearity of the inner product, we compute

$$(x \mid y) = \left(\sigma^x B \mid B' \sigma^{y\prime}\right) = \sigma^x \left(B \mid B'\right) \sigma^{y\prime}.$$

(b) The symmetry of the inner product implies that $(B \mid B')$ is a symmetric matrix. For any row vector $\alpha \in \mathbb{R}^d$, $\alpha (B \mid B') \alpha' = (\alpha B \mid \alpha B)$. By the positive definiteness of the inner product it follows that $\alpha (B \mid B') \alpha' \geq 0$, with equality holding if and only if $\alpha B = 0$. Since B is a basis, $\alpha B = 0$ if and only if $\alpha = 0$. This proves that $(B \mid B')$ is a positive definite matrix.

(c) By the linearity of the inner product, $x = \sigma^x B$ implies $(x \mid B') = \sigma^x (B \mid B')$. By part (b), $(B \mid B')$ is invertible, and the claimed expression for σ^x follows. ■

Two vectors x and y are **orthogonal** if $(x \mid y) = 0$. A set of vectors is **orthogonal** if its elements are pairwise orthogonal. A basis B is **orthonormal** if it is orthogonal and normalized so that $(B_i \mid B_i) = 1$ for every i, or, equivalently, if the Gram matrix $(B \mid B')$ is the identity matrix. Example 6 shows that every finite basis is orthonormal under some inner product. In Section 6 we will show that every finite-dimensional inner-product space has an orthonormal basis and can therefore be identified with a Euclidean inner-product space.

A **Riesz representation** of the linear functional $f : X \to \mathbb{R}$ is a vector $x \in X$ such that $f(y) = (x \mid y)$ for all $y \in X$.

Proposition 9 *Suppose $B = (B_1, \ldots, B_d)'$ is a basis of X and f is a linear functional. Then a Riesz representation of f exists, is unique and is given by*

$$f(B)' (B \mid B')^{-1} B.$$

Proof. For any $x = \sigma^x B$ and $y = \sigma^y B$,

$$f(y) = f(B)' \sigma^{y\prime} \quad \text{and} \quad (x \mid y) = \sigma^x (B \mid B') \sigma^{y\prime}.$$

The vector x is, therefore, a Riesz representation of f if and only if $f(B)' = \sigma^x (B \mid B')$. Since $(B \mid B')$ is positive definite, it is invertible, and therefore the last equality is equivalent to $\sigma^x = f(B)' (B \mid B')^{-1}$. ■

3 Norm

A **norm** on the vector space X is a function of the form $\|\cdot\| : X \to \mathbb{R}$ that satisfies the following properties for all $x, y \in X$:

1. (triangle inequality) $\|x + y\| \leq \|x\| + \|y\|$.
2. $\|\alpha x\| = |\alpha| \|x\|$ for all $\alpha \in \mathbb{R}$.

3. $\|x\| \geq 0$, with equality holding if and only if $x = 0$.

Here $\|x\|$ represents the value that the norm $\|\cdot\|$ assigns to x, referred to simply as the **norm of** x. It is easy to see that the triangle inequality holds for all $x, y \in X$ if and only if

$$|\|x\| - \|y\|| \leq \|x - y\| \quad \text{for all } x, y \in X. \tag{1}$$

In the current context, we assume that the norm $\|\cdot\|$ is **induced**[1] by the underlying inner product $(\cdot \mid \cdot)$, meaning that

$$\|x\| = \sqrt{(x \mid x)}, \quad x \in X. \tag{2}$$

We will see shortly that $\|\cdot\|$ so defined is indeed a norm. We think of $\|x\|$ as the length of x, in the sense used in our earlier informal motivation of inner products. Orthogonality of two vectors x and y, defined by the condition $(x \mid y) = 0$, can be characterized entirely in terms of the induced norm.

Proposition 10 *The vectors x and y are orthogonal if and only if they satisfy the **Pythagorean identity***

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

Proof. The claim follows from the computation

$$\|x + y\|^2 = (x + y \mid x + y) = (x \mid x) + (y \mid y) + 2\,(x \mid y) = \|x\|^2 + \|y\|^2 + 2\,(x \mid y) .$$

■

Two vectors x and y are said to be **colinear** if either $x = \alpha y$ or $y = \alpha x$ is true for some $\alpha \in \mathbb{R}$.

Proposition 11 *Equation* (2) *defines a norm $\|\cdot\|$ that satisfies the **Cauchy-Schwarz inequality***

$$|(x \mid y)| \leq \|x\| \, \|y\| , \quad x, y \in X,$$

with equality holding if and only if x and y are colinear.

[1] Every norm we use in this text is induced by some inner product. Not every norm can be induced by an inner product, however; a norm that is not translation invariant cannot be induced by an inner product.

Proof. The Cauchy-Schwarz inequality holds trivially as an equality if either x or y is zero. Suppose x and y are nonzero, and let $\hat{x} = x/\|x\|$ and $\hat{y} = y/\|y\|$. Visualizing the vector $(\hat{x} \mid \hat{y})\,\hat{y}$ as the projection of $\hat{x}$ on the line spanned by $\hat{y}$, we note that $\hat{x} - (\hat{x} \mid \hat{y})\,\hat{y}$ is orthogonal to $\hat{y}$. Indeed,

$$(\hat{x} - (\hat{x} \mid \hat{y})\,\hat{y} \mid \hat{y}) = (\hat{x} \mid \hat{y}) - (\hat{x} \mid \hat{y})\,(\hat{y} \mid \hat{y}) = 0.$$

The Pythagorean identity then implies that

$$0 \leq \|\hat{x} - (\hat{x} \mid \hat{y})\,\hat{y}\|^2 = 1 - (\hat{x} \mid \hat{y})^2\,,$$

which implies the Cauchy-Schwarz inequality. Equality holds if and only if $\hat{x} = (\hat{x} \mid \hat{y})\,\hat{y}$, a condition that is equivalent to $x = \alpha y$ for some scalar α. We still must verify that $\|\cdot\|$ is a norm. We use the Cauchy-Schwarz inequality to show the triangle inequality:

$$\begin{aligned} \|x + y\|^2 &= \|x\|^2 + 2\,(x \mid y) + \|y\|^2 \\ &\leq \|x\|^2 + 2\,\|x\|\,\|y\| + \|y\|^2 = (\|x\| + \|y\|)^2\,. \end{aligned}$$

The remaining norm properties are immediate from the definitions. ∎

4 Convergence

A sequence $\{x_n\} = (x_1, x_2, \ldots)$ of points in X **converges** to the **limit** $x \in X$ if for every $\varepsilon > 0$, there exists an integer N_ε such that $n > N_\varepsilon$ implies $\|x_n - x\| < \varepsilon$. In this case, the sequence $\{x_n\}$ is said to be **convergent.** A subset S of X is **closed** if every convergent sequence in S converges to a point in S. A subset S of X is **open** if its complement $X \setminus S$ is closed. The set of all open subsets of X is known as the **topology** of X.

The following properties of closed and open sets can be easily confirmed.[2] The empty set and X are both open and closed. The union of finitely many closed sets is closed, and the intersection of finitely many open sets is open. Arbitrary intersections of closed sets are closed, and arbitrary unions of open sets are open.

A sequence $\{x_n\}$ of points in X is **Cauchy** if for every $\varepsilon > 0$, there exists an integer N_ε such that $m, n > N_\varepsilon$ implies $\|x_m - x_n\| < \varepsilon$. A subset S of X is **complete** if every Cauchy

[2] Just as a norm need not be induced by an inner product, a topology, defined as a set of subsets of X with the stated properties, need not be defined by a norm in a more general theory of convergence, which is not needed in this text.

sequence in S converges to a limit in S. It is immediate from the definitions that a subset of a complete set is complete if and only if it is closed. The triangle inequality implies that every convergent sequence is Cauchy. Exercise 4 shows that the converse need not be true for an arbitrary inner product space. Intuitively, a Cauchy sequence should converge to something, but if that something is not within the space X, then the sequence is not convergent. As we will see shortly, difficulties of the sort do not arise in finite-dimensional spaces.

The following proposition shows that if X is finite dimensional, the convergence or Cauchy property of a sequence is equivalent to the respective property of the sequence's coordinates relative to any given basis.

Proposition 12 *Suppose $B = (B_1, \ldots, B_d)'$ is a basis of X, and $\sigma_n = \left(\sigma_n^1, \ldots, \sigma_n^d\right) \in \mathbb{R}^d$ for $n = 1, 2, \ldots$ The sequence $\{\sigma_n B\}$ is Cauchy (resp. converges to σB) if and only if the scalar sequence $\left\{\sigma_n^i\right\}$ is Cauchy (resp. converges to σ_i) for every coordinate $i \in \{1, \ldots, d\}$.*

Proof. Suppose $\left\{\sigma_n^i\right\}$ is Cauchy for each i. The triangle inequality implies $\|\sigma_n B - \sigma_m B\| \leq \sum_i \left|\sigma_n^i - \sigma_m^i\right| \|B_i\|$, from which it follows that $\{\sigma_n B\}$ is Cauchy. Conversely, suppose $\{\sigma_n B\}$ is Cauchy. Noting the identity $\sigma_n = (\sigma_n B \mid y)$, where $y = B' (B \mid B')^{-1}$, we apply the Cauchy-Schwarz inequality, to obtain $\left|\sigma_n^i - \sigma_m^i\right| = |(\sigma_n B - \sigma_m B \mid y_i)| \leq \|\sigma_n B - \sigma_m B\| \|y_i\|$. Therefore $\left\{\sigma_n^i\right\}$ is Cauchy. The claims in parentheses follow by the same argument, with σ in place of σ_m. ■

A **Hilbert space** is any inner product space that is complete (relative to the norm induced by the inner product). One of the fundamental properties of the real line is that it is complete. Given this fact, the last proposition implies:

Proposition 13 *Every finite-dimensional inner product space is a Hilbert space.*

Another consequence of Proposition 12 is that for a finite-dimensional vector space, the convergence or Cauchy property of a sequence does not depend on the choice of an underlying inner product, and therefore the topology on a finite-dimensional vector space is invariant to the inner product used to define it (or the norm used to define it, as explained in Section 8). The situation is radically different for infinite-dimensional spaces, where different norms can define different topologies.

We close this section with some topological terminology and notation that is of frequent use. Given any $r \in \mathbb{R}_+$ and $x \in X$, the **open ball** with center x and radius r is the set

$$B(x; r) = \{y \in X : \|y - x\| < r\}.$$

Let S be any subset of X. The **closure** of S is the set

$$\bar{S} = \{x : B(x; \varepsilon) \cap S \neq \emptyset \text{ for all } \varepsilon > 0\},$$

the **interior** of S is the set

$$S^0 = \{x : B(x; \varepsilon) \subset S \text{ for some } \varepsilon > 0\},$$

and the **boundary** of S is the set $\bar{S} \setminus S^0$. A vector x is a closure point of the set S if and only if there exists a sequence in S that converges to x. Therefore, the set S is closed if and only if $S = \bar{S}$, and it is open if and only if $S = S^0$. Finally, note that closures and interiors can be described in purely topological terms: The closure of a set is the intersection of all its closed supersets, and the interior of a set is the union of all its open subsets.

5 Continuity

A function $f : D \to \mathbb{R}$, where $D \subseteq X$, is **continuous at** $x \in D$ if for any sequence $\{x_n\}$ in D converging to x, the sequence $\{f(x_n)\}$ converges to $f(x)$. The function f is **continuous** if it is continuous at every point of its domain D. It is straightforward to check that $f : D \to \mathbb{R}$ is continuous at $x \in D$ if and only if given any $\varepsilon > 0$, there exists some $\delta > 0$ (depending on ε) such that $y \in D \cap B(x; \delta)$ implies $|f(y) - f(x)| < \varepsilon$. Based on last section's discussion, we note that if X is finite dimensional, the continuity of f at a point is true or not independently of the inner product (or norm) used to define the topology.

Inequality (1) shows that the underlying norm is a continuous function. The inner product $(\cdot \mid \cdot)$ is also continuous, in the following sense.

Proposition 14 *Suppose $\{x_n\}$ and $\{y_n\}$ are sequences in X converging to x and y, respectively. Then $\{(x_n \mid y_n)\}$ converges to $(x \mid y)$.*

Proof. By the triangle and Cauchy-Schwarz inequalities, we have

$$\begin{aligned} |(x_n \mid y_n) - (x \mid y)| &= |(x - x_n \mid y - y_n) + (x_n - x \mid y) + (x \mid y_n - y)| \\ &\leq \|x - x_n\| \|y - y_n\| + \|x_n - x\| \|y\| + \|x\| \|y_n - y\|. \end{aligned}$$

Letting n go to infinity completes the proof. ■

Exercise 5 gives an example of a linear functional on an infinite-dimensional space that is not continuous. In a finite-dimensional space, however, if f is a linear functional on X, then it has a Riesz representation z (by Proposition 9). An application of the Cauchy-Schwarz inequality shows that $|f(x) - f(y)| \leq \|z\| \|x - y\|$ for all $x, y \in X$ and hence the continuity of f. The remainder of this section extends this argument to concave functions, showing in particular that in a finite-dimensional space, a concave function over an open domain is continuous.

Consider any concave function $f : D \to \mathbb{R}$, where D is an open subset of X. We introduce two properties of f, which turn out to be equivalent to the continuity of f. We say that f is **locally bounded below** if given any $x \in D$, there exists a small enough radius $r > 0$ such that the infimum of f over the ball $B(x; r)$ is finite. We say that f is **locally Lipschitz continuous** if given any $x \in D$, there exist a constant K and a ball $B(x; r) \subseteq D$ such that

$$|f(y) - f(x)| \leq K \|y - x\| \quad \text{for all } y \in B(x; r). \tag{3}$$

Clearly, if f is locally Lipschitz continuous, it is continuous and therefore locally bounded below. The following lemma shows that the converse implications are also true.

Lemma 15 *Suppose that $D \subseteq X$ is open and convex, and the function $f : D \to \mathbb{R}$ is concave. If f is locally bounded below, then f is locally Lipschitz continuous.*

Proof. Given any $x \in D$, let b denote a lower bound of f on $B(x; r) \subseteq D$. Fixing any $y \in B(x; r)$, let $u = (y - x) / \|y - x\|$ and $\phi(\alpha) = f(x + \alpha u)$ for $\alpha \in [-r, r]$. Consider the following claim, where the first equality defines K :

$$\begin{aligned} K = \frac{\phi(0) - b}{r} &\geq \frac{\phi(0) - \phi(-r)}{r} \geq \frac{\phi(\|y - x\|) - \phi(0)}{\|y - x\|} \\ &\geq \frac{\phi(r) - \phi(0)}{r} \geq \frac{b - \phi(0)}{r} = -K. \end{aligned}$$

The function $\phi : [-r, r] \to \mathbb{R}$ is concave and bounded below by b, which justifies the first and last inequalities. The three middle expressions represent slopes that decrease from left to right since ϕ is concave. This proves (3). ■

Remark 16 *As in the lemma, assume that D is convex and open and $f : D \to \mathbb{R}$ is concave. Exercise 3 shows that f is locally bounded below if it is bounded below over some open ball within its domain. A corollary of this is that f is locally Lipschitz continuous if it is continuous at a single point of its domain.*

The preceding lemma (and remark) applies to an arbitrary inner product space X. Further restricting X to be finite-dimensional, we can verify that a concave function is locally bounded below, thus proving its continuity.

Theorem 17 *Suppose that X is finite dimensional, D is an open convex subset of X and $f : D \to \mathbb{R}$ is concave. Then f is locally Lipschitz continuous, and therefore continuous.*

Proof. Suppose first that $X = \mathbb{R}^d$ with the standard Euclidean inner product. We fix any $x \in D$ and show that f is bounded below near x. An exercise shows that x is contained in the interior of a set of the form

$$[\alpha, \beta] = \{x \in X : x_i \in [\alpha_i, \beta_i]\,,\ i = 1, \ldots, d\} \subseteq D,$$

where $\alpha, \beta \in \mathbb{R}^d$ and $\alpha_i < \beta_i$ for all i. Since f is concave, it is minimized over $[\alpha, \beta]$ at some **extreme** point of $[\alpha, \beta]$, that is, a point x such that $x_i \in \{\alpha_i, \beta_i\}$ for all i. To see why, take any $x \in [\alpha, \beta]$ and any $k \in \{1, \ldots, d\}$, and define the points x^α and x^β by $x_i^\alpha = x_i = x_i^\beta$ for $i \neq k$, $x_k^\alpha = \alpha_k$ and $x_k^\beta = \beta_k$. Then for some $\rho \in [0, 1]$, $x = \rho x^\alpha + (1 - \rho)\, x^\beta$. Concavity implies $\min\left\{f(x^\alpha), f\left(x^\beta\right)\right\} \leq f(x)$. We can therefore replace x by one of x^α or x^β without increasing its value under f. Repeating this process for all coordinates shows that for every $x \in [\alpha, \beta]$, there exists some extreme point $\bar{x}$ of $[\alpha, \beta]$ such that $f(x) \geq f(\bar{x})$. Since $[\alpha, \beta]$ has only finitely many extreme points, f is bounded below on $[\alpha, \beta]$. Continuity of f at x follows from Lemma 15.

Finally, suppose X is an arbitrary finite-dimensional space and let B be any basis. We have seen in Proposition 12 that whether f is continuous or not does not depend on what inner product we endow X with. We can therefore choose the inner product of Example 6, which makes B an orthonormal basis. Identifying each element $x = \sigma^x B \in X$ with its basis representation σ^x, the above argument applies. ■

6 Projections on Convex Sets

Projections were introduced informally in motivating inner products. This section defines projections formally, shows the existence of projections on complete convex sets, and establishes a dual characterization of projections on convex sets in terms of inner products.

Definition 18 *Let S be any subset of X. The vector x_S is a* ***projection*** *of x on the set S if $x_S \in S$ and $\|x - s\| \geq \|x - x_S\|$ for all $s \in S$. The vector $z \in X$* ***supports*** *the set S at $\bar{s} \in S$ if $(z \mid s - \bar{s}) \geq 0$ for all $s \in S$.*

The inequality $(z \mid s - \bar{s}) \geq 0$ can be visualized as the requirement that the vectors z and $s - \bar{s}$ form an acute angle. Theorem 19(a) below characterizes projections on a convex set S based on the intuitive geometric idea that a point x_S in S is the projection on S of a point x outside S if and only if for any $s \in S$ the vectors $x_S - x$ and $s - x_S$ form an acute angle. The theorem also addresses the uniqueness and existence of a projection on a convex set. Clearly, if S is not closed then a projection of a point on S may not exist. For example, on the real line, the projection of zero on the open interval $(1, 2)$ does not exist. Exercise 4 further shows that in an infinite-dimensional space, S can be closed and convex, and yet the projection of zero on S may still not exist. The key issue in this case is the absence of completeness of the ambient space X.

The central result on projections on convex sets follows. The theorem and its proof apply to any inner product space (not necessarily a finite-dimensional one).

Theorem 19 (Projection Theorem) *Suppose that S is a convex subset of X and $x, y \in X$.*

(a) The vector x_S is a projection of x on S if and only if $x_S \in S$ and $x_S - x$ supports S at x_S.

(b) Suppose x_S is a projection of x on S and y_S is a projection of y on S. Then $\|x_S - y_S\| \leq \|x - y\|$.

(c) There exists at most one projection of x on S.

(d) If S is complete, then the projection of x on S exists.

Proof. (a) Suppose $x_S - x$ supports S at $x_S \in S$. Then for any $s \in S$,

$$\|x - s\|^2 = \|x - x_S\|^2 + \|s - x_S\|^2 + 2\,(x_S - x \mid s - x_S) \geq \|x - x_S\|^2 ,$$

proving that x_S is a projection of x on S. Conversely, suppose x_S is a projection of x on S. Given any nonzero $s \in S$ and $\alpha \in (0, 1)$, define $x_\alpha = x_S + \alpha\,(s - x_S) \in S$. Then

$$\begin{aligned} 2\,(x_S - x \mid s - x_S) &= -\alpha\,\|s - x_S\|^2 + \frac{1}{\alpha}\left(\|x - x_\alpha\|^2 - \|x - x_S\|^2\right) \\ &\geq -\alpha\,\|s - x_S\|^2 . \end{aligned}$$

Letting α approach zero, we conclude that $(x_S - x \mid s - x_S) \geq 0$.

(b) Let $\delta = y - x$ and $\delta_S = y_S - x_S$. By part (a), $(x_S - x \mid \delta_S) \geq 0$ and $(y - y_S \mid \delta_S) \geq 0$. Adding the two inequalities, we obtain $(\delta - \delta_S \mid \delta_S) \geq 0$. Finally,

$$\|\delta\|^2 = \|\delta - \delta_S\|^2 + \|\delta_S\|^2 + 2\,(\delta - \delta_S \mid \delta_S) \geq \|\delta_S\|^2 .$$

(c) Apply part (b) with $x = y$.

(d) Let $\delta = \inf\{\|x - s\| : s \in S\}$, and choose a sequence $\{s_n\}$ in S such that $\{\|x - s_n\|\}$ converges to δ. Direct computation (see Exercise 2) shows

$$\begin{aligned}\|s_m - s_n\|^2 &= 2\|x - s_m\|^2 + 2\|x - s_n\|^2 - 4\left\|x - \frac{s_m + s_n}{2}\right\|^2 \\ &\leq 2\|x - s_m\|^2 + 2\|x - s_n\|^2 - 4\delta^2,\end{aligned}$$

implying $\|s_m - s_n\|^2 \to 0$ as $m, n \to \infty$. Therefore, the sequence $\{s_m\}$, being Cauchy, converges to some vector x_S in S, since S is assumed complete. By continuity of the norm operator it follows that $\delta = \|x - x_S\|$ and therefore x_S is a projection of x on S. ■

Corollary 20 *If X is a Hilbert space and S is a closed convex subset of X, the projection of any $x \in X$ on S exists.*

7 Orthogonal Projections

This section elaborates on the important special case of projections on linear manifolds. A vector x is **orthogonal** to the linear manifold M if x is orthogonal to $m_1 - m_2$ for all $m_1, m_2 \in M$. The **orthogonal to M subspace**, denoted by $M^{\perp}$, is the linear subspace of all vectors that are orthogonal to M. Note that $M^{\perp} = (x + M)^{\perp}$ for every $x \in X$, and a vector supports M at some point if and only if it is orthogonal to M. Theorem 19(a) applied to linear manifolds reduces to the following result.

Corollary 21 (orthogonal projections) *Suppose M is a linear manifold in X and $x \in X$. A point x_M is the projection of x on M if and only if $x_M \in M$ and $x - x_M \in M^{\perp}$.*

Projections on finite-dimensional subspaces can be expressed by simple formulas in terms of a given basis. Note that the following proposition, which generalizes Proposition 8(c), does not assume that the ambient space X is finite dimensional.

Proposition 22 *Suppose L is a linear subspace of X, and $B = (B_1, \ldots, B_d)'$ is a basis of L. Given any $x \in X$, the vector*

$$x_L = \left(x \mid B'\right)\left(B \mid B'\right)^{-1} B \tag{4}$$

is the projection of x on L, and $x - x_L$ is the projection of x on

$$L^{\perp} = \{y \in X : (B \mid y) = 0\}. \tag{5}$$

Finally, $L^{\perp\perp} = L$.

Proof. Let $x_L = \sigma^{x_L} B$, where $\sigma^{x_L} = (x \mid B')(B \mid B')^{-1}$. Clearly, $x_L \in L$. Moreover, $x - x_L \in L^{\perp}$, since, for any $y = \sigma^y B \in L$,

$$(x_L \mid y) = \sigma^{x_L} \left(B \mid B'\right) \sigma^{y\prime} = \left(x \mid B'\right) \left(B \mid B'\right)^{-1} \left(B \mid B'\right) \sigma^{y\prime} = (x \mid y) \,.$$

By Corollary 21, x_L is the projection of x on L. It is immediate that $L \subseteq L^{\perp\perp}$. Therefore, $x - (x - x_L) = x_L \in L^{\perp\perp}$ and $x - x_L \in L^{\perp}$. Again by Corollary 21, it follows that $x - x_L$ is the projection of x on $L^{\perp}$. Finally, we show that $L^{\perp\perp} \subseteq L$. Consider any $x \in L^{\perp\perp}$ and let x_L be its projection on L, given by (4). We can then write $x = x_L + n$, where $n \in L^{\perp}$. Since $(x \mid n) = (x_L \mid n) + (n \mid n)$ and $(x \mid n) = (x_L \mid n) = 0$, it follows that $(n \mid n) = 0$ and therefore $x = x_L \in L$. ∎

Corollary 23 *Suppose that* $B = (B_1, \ldots, B_d)'$ *is a column matrix of linearly independent vectors in* X, *and* $M = \{y \in X : (B \mid y) = b\}$ *for some column vector* $b \in \mathbb{R}^d$. *Then* $M^{\perp} = \operatorname{span}(B_1, \ldots, B_d)$, *and the projection of* x *on* M *exists and is given by*

$$x_M = x - \left(\left(x \mid B'\right) - b'\right) \left(B \mid B'\right)^{-1} B.$$

Proof. Fixing any $m \in M$, note that $M - m$ is the linear subspace $L^{\perp}$ of equation (5), and therefore $M^{\perp} = L^{\perp\perp} = L = \operatorname{span}(B)$. The point x_M is the projection of x on M if and only if $x_M - m$ is the projection of $x - m$ on $L^{\perp}$. We can therefore apply Proposition (22) to conclude that the projection x_M of x on M exists and

$$x_M - m = (x - m) - \left(x - m \mid B'\right) \left(B \mid B'\right)^{-1} B.$$

Since $(B \mid m) = b$, the claimed expression for x_M follows. ∎

Corollary 24 *Suppose* M *is a finite-dimensional linear manifold in* X. *Then every vector* $x \in X$ *has a unique decomposition of the form* $x = m + n$, *where* $m \in M$ *and* $n \in M^{\perp}$.

The preceding corollary is not valid if the assumption that M is finite-dimensional is relaxed, in which case M need not be closed. The result extends easily to any Hilbert space, however, still as a corollary of the Projection Theorem, provided the vector m is only required to lie in the closure of M (rather than M).

If the linear subspace L has a finite orthogonal basis B, then $(B \mid B')$ is diagonal and formula (4) for the projection of x on L reduces to

$$x_L = \sum_{i=1}^{k} \frac{(x \mid B_i)}{(B_i \mid B_i)} B_i. \tag{6}$$

Assuming X is finite-dimensional, this equation can be used to recursively construct an orthogonal basis of X, a process known as **Gram-Schmidt orthogonalization**. We start with any nonzero vector B_1 in X. Assuming we have constructed k orthogonal vectors $B_1, \ldots, B_k$ such that $L = \text{span}(B_1, \ldots, B_k) \neq X$, we select any $x \in X \setminus L$ and define $B_{k+1} = x - x_L$, where x_L is the projection of x on L, given by (6). If X is d-dimensional, the recursive construction terminates for $k = d$. Normalizing the resulting orthogonal basis proves

Proposition 25 *Every finite-dimensional inner product space has an orthonormal basis.*

Formula (4) is the same as the Riesz representation expression of Proposition 9 with $f(y) = (x \mid y)$. The reason should be clear given the following general relationship between orthogonal projections and Riesz representations (whose proof is immediate from the definitions).

Proposition 26 *Suppose that $f(y) = (x \mid y)$ for all $y \in X$ and f_L is the restriction of f on the linear subspace L. The vector x_L is the Riesz representation of f_L in L if and only if it is the projection of x on L.*

We use orthogonal projections to show that in any Hilbert space the Riesz representation of a continuous linear functional exists. The argument is redundant in the finite-dimensional case, since the claim was established in Proposition 9, but still worth reviewing.

Theorem 27 *In a Hilbert space, a linear functional is continuous if and only if it has a (necessarily unique) Riesz representation.*

Proof. Suppose f is a continuous linear functional. The null subspace $N = \{x : f(x) = 0\}$ is closed. If $X = N$, the result follows trivially. Otherwise there exists $z \in X$ such that $f(z) \neq 0$. Let z_N be the projection of z on N, and define $y = f(z - z_N)^{-1}(z - z_N)$. Then y is orthogonal to N and satisfies $f(y) = 1$. For any $x \in X$, the fact that $x - f(x)\, y \in N$ implies that $(x - f(x)\, y \mid y) = 0$, and therefore $f(x) = (x \mid y) / (y \mid y)$. The normalized vector $(y \mid y)^{-1} y$ is the Riesz representation of f. ■

8 Compactness

In this section we show that a continuous function achieves a maximum and a minimum over any compact set. While the result applies to any normed space, it is mostly useful in the finite-dimensional case, in which a set is compact if and only if it is closed and bounded.

A subset S of X is (sequentially) **compact** if every sequence in S has a subsequence that converges to a vector in S. This definition immediately implies that a compact set is complete and therefore closed. A compact set S is also **bounded**, meaning that $\sup\{\|s\| : s \in S\} < \infty$. (If S were unbounded, there would exist a sequence $\{s_n\}$ in S such that $\|s_n\| > n$ for all n, a condition that precludes the existence of a convergent subsequence.)

Proposition 28 *In a finite-dimensional space, a set is compact if and only if it is closed and bounded.*

Proof. As noted above, it is generally true that a compact set is closed and bounded. We prove the converse, relying on the assumption that X is finite dimensional. By Proposition 12 and Example 6, it suffices to show the result for $X = \mathbb{R}^d$ (why?). Let $\{s_n\} = \left\{\left(s_n^1, \ldots, s_n^d\right)\right\}$ be a sequence in a closed bounded subset S of $\mathbb{R}^d$. Then the first coordinate sequence $\left\{s_n^1\right\}$ lies in some bounded interval I. Select a half-interval I_1 of I that contains infinitely many points of $\left\{s_n^1\right\}$ and let $s_{n_1}^1$ be one of these points. Then select a half-interval I^2 of I^1 that contains infinitely many points of $\left\{s_n^1 : n > n_1\right\}$ and let $s_{n_2}^1$ be one of these points. Continuing in this manner, we obtain a nested sequence of intervals $\{I_k\}$ whose length shrinks to zero and a corresponding subsequence $\left\{s_{n_k}^1\right\}$ with $s_{n_k}^1 \in I_k$ for all k. Clearly, the subsequence $\left\{s_{n_k}^1\right\}$ is Cauchy and therefore convergent. Repeating the argument we can extract a further subsequence for the second coordinate, then the third, and so on. This process generates a convergent subsequence of $\{s_n\}$ whose limit point must be in S, since S is assumed closed. ■

The preceding proposition is not valid in infinite-dimensional spaces. It is not hard to prove that a set S in any normed space is compact if and only if it is complete and for every $\varepsilon > 0$ there exists a finite collection of balls of radius ε whose union contains S. This generalization will not be useful in our applications, however—a more fruitful direction is briefly discussed at the end of this section.

Proposition 29 *Suppose that S is a compact subset of X and the function $f : S \to \mathbb{R}$ is continuous. Then there exist $s^*, s_* \in S$ such that*

$$f(s^*) \geq f(s) \geq f(s_*) \quad \text{for all } s \in S.$$

Proof. Let $\{s_n\}$ be a sequence such that $\lim_n f(s_n) = \sup f$. By the compactness of S, there exists a subsequence of $\{s_n\}$ converging to some $s^* \in S$. Since f is continuous, $f(s^*) = \sup f$ and therefore $f(s^*) \geq f(s)$ for all $s \in S$. The same argument applied to $-f$ completes the proof. ■

Proposition 30 *Suppose X is finite-dimensional, A and B are closed subsets of X and at least one of them is bounded. Then there exists a pair $(a, b) \in A \times B$ such that $\|a - b\| \leq \|x - y\|$ for all $(x, y) \in A \times B$.*

Proof. Suppose A is bounded and therefore compact. Select $r > 0$ large enough so that the set $C = B \cap \{y : \|x - y\| \leq r \text{ for some } x \in A\}$ is nonempty. Note that C is also compact (why?). Let $\{(a_n, b_n)\}$ be any sequence in $A \times C$ such that $\|a_n - b_n\|$ converges to $\delta = \inf \{\|x - y\| : (x, y) \in A \times C\}$. We can then extract a subsequence $\{(a_{n_k}, b_{n_k})\}$ that converges to some $(a, b) \in A \times C$. By the triangle inequality,

$$\|a - b\| \leq \|a - a_{n_k}\| + \|a_{n_k} - b_{n_k}\| + \|b_{n_k} - b\|.$$

The right-hand side converges to δ, and therefore $\|a - b\| = \delta$. For any $(x, y) \in A \times B$, either $y \in C$, in which case $\|x - y\| \geq \delta$ by the definition of δ, or $y \notin C$, in which case $\|x - y\| > r$ by the definition of C. Since $r \geq \delta$, the proof is complete. ■

Proposition 29 implies the equivalence of all norms on a finite-dimensional space, in the following sense. Suppose X is finite dimensional and consider any other norm $\|\cdot\|_*$ on X (not necessarily induced by any inner product). Since $\|\cdot\|_*$ is convex, it is continuous on the open ball $B(0; 2)$ and therefore achieves a minimum and a maximum over the closure of $B(0; 1)$. By the homogeneity of $\|\cdot\|_*$, there exist constants k and K such that $k\|x\| \leq \|x\|_* \leq K\|x\|$ for all $x \in X$. Therefore, for a finite-dimensional vector space, all norms define the same topology and the same compact sets.

In infinite-dimensional spaces, closed and bounded sets are not compact and different norms can define dramatically different topologies. This severely limits the usefulness of compactness, in this section's sense, for infinite-dimensional optimization theory. A partial remedy is provided by the notion of weak compactness, which is particularly useful in conjunction with convexity properties. For example, a weak compactness argument can be used to prove that in any Hilbert space, if the nonempty set S is convex, closed and bounded, and the function $f : X \to \mathbb{R}$ is concave over S and continuous at every point of S, then f achieves a maximum in S. Suitable references are provided in the endnotes.

9 Supporting Hyperplanes

A **hyperplane** H is a linear manifold whose orthogonal subspace is of dimension one. If $H^{\perp}$ is spanned by the vector y and $\alpha = (y \mid \bar{x})$, where $\bar{x}$ is any point in H, then

$$H = \{x : (y \mid x) = \alpha\}. \tag{7}$$

This expression characterizes all hyperplanes as y ranges over the set of all nonzero vectors and α ranges over the set of all scalars.

According to Definition 18, the vector y supports the set $S \subseteq X$ at $\bar{s} \in S$ if and only if $(y \mid \bar{s}) \leq (y \mid s)$ for all $s \in S$. Letting $\alpha = (y \mid \bar{s})$ in (7), this condition can be visualized as S being included in the half-space $\{x : (y \mid x) \geq \alpha\}$ on the one side of the hyperplane H, while touching H at $\bar{s}$. The following is an extension of Definition 18, which can be visualized the same way, but does not require $\bar{s}$ to be an element of S.

Definition 31 *The vector $y \in X$ **supports** the set S at $\bar{s} \in X$ if*

$$(y \mid \bar{s}) = \inf \{(y \mid s) : s \in S\}. \tag{8}$$

It is intuitively compelling that one should be able to support a convex set at any point of their boundary. The supporting hyperplane theorem, proved below, shows that this is indeed true for a finite-dimensional space. In another example in which finite-dimensional intuition need not apply in infinite-dimensional spaces, Exercise 16 shows that in an infinite-dimensional space it may not be possible to support a convex set at a point of its boundary.[3]

Theorem 32 (Supporting Hyperplane Theorem) *Suppose S is a convex subset of the finite-dimensional inner-product space X. Then, given any vector x that is not in the interior of S, there exists a nonzero vector y such that $(y \mid x) \leq (y \mid s)$ for all $s \in S$. If x is on the boundary of S, then y supports S at x.*

[3] The supporting hyperplane theorem does stay true if the assumption of a finite-dimensional X is replaced by the assumption that S has an interior point. But, as in the example of Exercise 16, this is typically too restrictive an assumption for applications.

Proof. Since x is not interior, we can construct a sequence $\{x_n\}$ of vectors such that $x_n \notin \bar{S}$ for all n and $x_n \to x$ as $n \to \infty$. For each n, let $\bar{s}_n$ be the projection of x_n on $\bar{S}$ and define $y_n = (\bar{s}_n - x_n) / \|\bar{s}_n - x_n\|$. The dual characterization of projections gives $(y_n \mid \bar{s}_n) \leq (y_n \mid s)$ for all $s \in S$. By Theorem 19(b), the sequence $\{\bar{s}_n\}$ converges to the projection $\bar{s}$ of x on $\bar{S}$. The sequence $\{y_n\}$ lies in the closure of the unit ball, which is compact, and therefore we can extract a subsequence $\{y_{n_k}\}$ that converges to some vector y of unit norm. By the continuity of inner products, we conclude that $(y \mid \bar{s}) \leq (y \mid s)$ for all $s \in S$. If x is on the boundary of S, then $x = \bar{s}$ is the limit of some sequence $\{s_n\}$ in S. Therefore, $\{(y \mid s_n)\}$ converges to $(y \mid \bar{s})$, implying (8). If x is not on the boundary of S, then $\{y_{n_k}\}$ converges to $y = (\bar{s} - x) / \|\bar{s} - x\|$. Since $(y \mid \bar{s} - x) > 0$, it follows that $(y \mid x) < (y \mid \bar{s}) \leq (y \mid s)$ for all $s \in S$. ■

Corollary 33 (Separating Hyperplane Theorem) *Suppose A and B are convex subsets of the finite-dimensional inner-product space X. If $A \cap B = \emptyset$, then there exists a nonzero $y \in X$ such that*

$$\inf_{a \in A} (y \mid a) \geq \sup_{b \in B} (y \mid b). \tag{9}$$

Proof. Let $S = A - B = \{a - b : a \in A,\ b \in B\}$. Since $A \cap B = \emptyset$, zero is not in S. If X is infinite-dimensional, the fact that at least one of A and B has a nonempty interior implies that S has a nonempty interior. By the supporting hyperplane theorem, there exists a nonzero $y \in X$ such that $(y \mid s) \geq 0$ for all $s \in S$ and therefore $(y \mid a) \geq (y \mid b)$ for every $a \in A$ and $b \in B$. ■

10 Superdifferential and Gradient

Consider any function $f : D \to \mathbb{R}$, where $D \subseteq X$. The vector $y \in X$ is a **supergradient** of f at $x \in D$ if it satisfies the **gradient inequality**:

$$f(x+h) \leq f(x) + (y \mid h) \text{ for all } h \text{ such that } x + h \in D.$$

The **superdifferential** of f at x, denoted by $\partial f(x)$, is the set of all supergradients of f at x.

The supergradient property can be visualized as a support condition in the space $X \times \mathbb{R}$ with the inner product

$$((x_1, \alpha_1) \mid (x_2, \alpha_2)) = (x_1 \mid x_2) + \alpha_1 \alpha_2, \quad x_i \in X,\ \alpha_i \in \mathbb{R}.$$

The **subgraph** of f is the set

$$\operatorname{sub}(f) = \{(x, \alpha) \in D \times \mathbb{R} : \alpha \leq f(x)\}. \tag{10}$$

The preceding definitions imply that

$$y \in \partial f(x) \quad \Longleftrightarrow \quad (y, -1) \text{ supports } \operatorname{sub}(f) \text{ at } (x, f(x)). \tag{11}$$

Theorem 34 *In a finite-dimensional inner product space,*[4] *the superdifferential of a concave function at an interior point of its domain is nonempty, convex and compact.*

Proof. Suppose $f : D \to \mathbb{R}$ is concave and $x \in D^0$. By the supporting hyperplane theorem (Theorem 32), sub(f) is supported by some nonzero $(y, -\beta) \in X \times \mathbb{R}$ at $(x, f(x))$:

$$(y \mid x) - \beta f(x) = \min\{(y \mid z) - \beta\alpha : \alpha \leq f(z),\ z \in D,\ \alpha \in \mathbb{R}\}.$$

Since the left-hand side is finite, it follows that $\beta \geq 0$. If $\beta = 0$, then y supports D at x, which contradicts the assumption that x is an interior point of D. Therefore, $\beta > 0$ and $y/\beta \in \partial f(x)$. This proves that $\partial f(x)$ is nonempty. It follows easily from the definitions that $\partial f(x)$ is also convex and closed. Finally, we show that $\partial f(x)$ is bounded, utilizing the finite dimensionality assumption. We assume $x = 0$, which entails no loss of generality (why?). Theorem 17 implies that f is bounded below over some ball centered at zero. Let $\varepsilon > 0$ and $K \in \mathbb{R}$ be such that for all $z \in X$, $\|z\| = \varepsilon$ implies $z \in D^0$ and $f(z) > K$. It follows that for every $y \in \partial f(0)$,

$$K < f\left(-\frac{\varepsilon}{\|y\|}y\right) \leq f(0) + \left(y \mid -\frac{\varepsilon}{\|y\|}y\right) = f(0) - \varepsilon\|y\|,$$

which results in the bound $\|y\| < (f(0) - K)/\varepsilon$. This proves that $\partial f(x)$ bounded. Since it is also closed, it is compact. ■

Assuming it exists, the **directional derivative of f at x in the direction h** is denoted and defined by

$$f'(x; h) = \lim_{\alpha \downarrow 0} \frac{f(x + \alpha h) - f(x)}{\alpha}.$$

The gradient of f at x is said to exist if $f'(x; h)$ exists for every $h \in X$ and the functional $f'(x; \cdot)$ is linear. In this case, the Riesz representation of the linear functional $f'(x; \cdot)$ is

[4] The result extends to general Hilbert spaces, provided the compactness conclusion is replaced by weak compactness.

the **gradient** of f at x, denoted by $\nabla f(x)$. Therefore, when it exists, the gradient $\nabla f(x)$ is characterized by the restriction

$$f'(x;h) = (\nabla f(x) \mid h), \quad \text{for all } h \in X.$$

For concave functions, directional derivatives and gradients are related to superdifferentials as follows.

Theorem 35 *Suppose D is a convex subset of the finite-dimensional inner product space X, the function $f : D \to \mathbb{R}$ is concave, and x is any interior point of D. Then the directional derivative $f'(x;h)$ exists for all $h \in X$, and the following three claims are true:*

(a) $\partial f(x) = \{y \in X \mid (y \mid h) \geq f'(x;h) \text{ for all } h \in X\}$.

(b) $f'(x;h) = \min\{(y \mid h) : y \in \partial f(x)\}$ for every $h \in X$.

(c) The gradient $\nabla f(x)$ exists if and only if $\partial f(x)$ is a singleton, in which case $\partial f(x) = \{\nabla f(x)\}$.

Proof. Fixing any $h \in X$, let $A = \left\{\alpha \in \mathbb{R} : x + \alpha h \in D^0\right\}$ (an open interval) and define the concave function $\phi : A \to \mathbb{R}$ by $\phi(\alpha) = f(x + \alpha h) - f(x)$. Note that $f'(x;h) = \phi'_+(0)$ (the right derivative ϕ at zero). For each $\alpha \in A$, let $\Delta(\alpha) = \phi(\alpha)/\alpha$, which is the slope of the line segment on the plane connecting the origin to the point $(\alpha, \phi(\alpha))$. Consider any decreasing sequence $\{\alpha_n\}$ of positive scalars in A. Concavity of ϕ implies that the corresponding sequence of slopes $\{\Delta(\alpha_n)\}$ is increasing. Moreover, fixing any $\varepsilon > 0$ such that $x - \varepsilon h \in D^0$, we have $\Delta(\alpha_n) \leq \Delta(-\varepsilon)$ for all n. Being increasing and bounded, the sequence $\{\Delta(\alpha_n)\}$ converges. This proves that the (monotone) limit $f'(x;h) = \lim_{\alpha \downarrow 0} \Delta(\alpha)$ exists and is finite.

(a) If $y \in \partial f(x)$, the gradient inequality implies that $(y \mid h) \geq \Delta(\alpha)$ for all positive $\alpha \in A$. Letting $\alpha \downarrow 0$, it follows that $(y \mid h) \geq f'(x;h)$. Conversely, suppose that $y \notin \partial f(x)$, and therefore $\Delta(1) = f(x+h) - f(x) > (y \mid h)$ for some h. We saw earlier that $\Delta(\alpha)$ increases to $f'(x;h)$ as $\alpha \downarrow 0$, and therefore $f'(x;h) \geq \Delta(1) > (y \mid h)$ for some h.

(b) In light of part (a), the claim of part (b) follows if given any $h \in X$, we can produce some $y \in \partial f(x)$ such that $(y \mid h) = f'(x;h)$. The gradient inequality $\phi(\alpha) \leq \phi'_+(0)\,\alpha$ for all $\alpha \in A$ (equivalently, $f(x + \alpha h) \leq f(x) + \alpha f'(x;h)$ for all $\alpha \in A$), can be restated as the condition that the line segment

$$L = \left\{\left(x + \alpha h, f(x) + \alpha f'(x;h)\right) : \alpha \in A\right\} \subseteq X \times \mathbb{R}$$

does not intersect the interior of the subgraph of f, as defined in (10). By the separating hyperplane theorem (Corollary 33), there exists nonzero $(p, \beta) \in X \times \mathbb{R}$ that separates the sets L and $\mathrm{sub}(f)$:

$$\inf_{\alpha \in A} (p \mid x + \alpha h) + \beta \left(f(x) + \alpha f'(x; h) \right) \geq \sup_{(\tilde{x}, \alpha) \in \mathrm{sub}(f)} (p \mid \tilde{x}) + \beta \alpha.$$

Since the right-hand side must be finite, $\beta > 0$. It follows that the right-hand side is at least as large as $(p \mid x) + \beta f(x)$, which is also obtained as the expression on the left-hand side with $\alpha = 0$. Since $0 \in A^0$, the coefficient of α on the left-hand side must vanish: $(p \mid h) + \beta f'(x; h) = 0$. Therefore, $f'(x; h) = (y \mid h)$, where $y = -p/\beta$, and the separation condition reduces to the gradient inequality defining the condition $y \in \partial f(x)$.

(c) If $\partial f(x) = \{\delta\}$, then part (b) implies that $f'(x; h) = (\delta \mid h)$ for all $h \in X$, and therefore $\delta = \nabla f(x)$. Conversely, if the gradient exists, then part (a) implies that $(\nabla f(x) - y \mid h) \leq 0$ for all $h \in X$ and $y \in \partial f(x)$. Letting $h = \nabla f(x) - y$, it follows that $y = \nabla f(x)$ if $y \in \partial f(x)$. ■

11 Global Optimality Conditions

We take as given the set $C \subseteq X$, and the functions $U : C \to \mathbb{R}$ and $G : C \to \mathbb{R}^n$, for some positive integer n. For each $\delta \in \mathbb{R}^n$, we consider the constrained optimization problem

$$J(\delta) = \sup \{U(x) : G(x) \leq \delta, \ x \in C\}. \qquad (P_\delta)$$

With the convention $\inf \emptyset = -\infty$, this defines a function of the form $J : \mathbb{R}^n \to [-\infty, +\infty]$, which is clearly monotone: $\delta_1 \geq \delta_2 \implies J(\delta_1) \geq J(\delta_2)$. We will characterize the solution to the problem (P_0). Since problem (P_δ) is the same as problem (P_0) with $G - \delta$ in place of G, this covers the general case. The key to understanding (P_0), however, is to consider the entire function J.

Associated with problem (P_0) is the **Lagrangian**

$$\mathcal{L}(x, \lambda) = U(x) - \lambda \cdot G(x), \quad x \in C, \quad \lambda \in \mathbb{R}^n,$$

where the dot denotes the Euclidean inner product in $\mathbb{R}^n$. The parameter λ will be referred to as a **Lagrange multiplier**. Assuming $J(0)$ is finite, we extend our earlier definition (for finite-valued functions) by defining the **superdifferential** of J at zero to be the set

$$\partial J(0) = \{\lambda \in \mathbb{R}^n : J(\delta) - J(0) \leq \lambda \cdot \delta \text{ for all } \delta \in \mathbb{R}^n\}.$$

The monotonicity of J implies that $\partial J(0) \subseteq \mathbb{R}^n_+$. Indeed, if $\lambda \in \partial J(0)$, then $0 \leq J(\delta) - J(0) \leq \lambda \cdot \delta$ for all $\delta \geq 0$ and therefore $\lambda \geq 0$. The following relationship between the Lagrangian and the superdifferential of J at zero is key.

Lemma 36 *Suppose that $J(0)$ is finite. Then for every $\lambda \in \mathbb{R}^n_+$,*

$$\lambda \in \partial J(0) \qquad \Longleftrightarrow \qquad J(0) = \sup\{\mathcal{L}(x, \lambda) : x \in C\}.$$

Proof. In the space $\mathbb{R}^n \times \mathbb{R}$ with the Euclidean inner product, $\lambda \in \partial J(0)$ if and only if $(\lambda, -1)$ supports at $(0, J(0))$ (in the sense of Definition 31) the set

$$S_1 = \{(\delta, \upsilon) \in \mathbb{R}^n \times \mathbb{R} : \upsilon < J(\delta)\}.$$

Similarly, $J(0) = \sup\{\mathcal{L}(x, \lambda) : x \in C\}$ if and only if $(\lambda, -1)$ supports at $(0, J(0))$ the set

$$S_2 = \{(\delta, \upsilon) \in \mathbb{R}^n \times \mathbb{R} : G(x) \leq \delta \text{ and } \upsilon < U(x) \text{ for some } x \in C\}.$$

The proof is concluded by noting that $S_1 = S_2$. ■

Lemma 36 does not assume that an optimum is achieved. If a maximum does exist, the argument can be extended to obtain the following global optimality conditions for problem (P_0).

Theorem 37 *Suppose that $c \in C$, $G(c) \leq 0$ and $\lambda \in \mathbb{R}^n$. Then the following two conditions are equivalent:*

1. *$U(c) = J(0)$ and $\lambda \in \partial J(0)$.*
2. *$\mathcal{L}(c, \lambda) = \max_{x \in C} \mathcal{L}(x, \lambda), \quad \lambda \cdot G(c) = 0, \quad \lambda \geq 0$.*

Proof. $(1 \implies 2)$ Suppose condition 1 holds. We noted that $\lambda \in \partial J(0)$ implies $\lambda \geq 0$. By the last lemma, $U(c) = \sup\{\mathcal{L}(x, \lambda) : x \in C\}$. Since $G(c) \leq 0$, we also have $\mathcal{L}(c, \lambda) \geq U(c)$, and therefore the last inequality is an equality, proving that $\lambda \cdot G(c) = 0$.

$(2 \implies 1)$ Condition 2 implies that $U(c) = \mathcal{L}(c, \lambda) \geq \mathcal{L}(x, \lambda) \geq U(x)$ for every x such that $G(x) \leq 0$. Therefore, $U(c) = J(0)$, and condition 1 follows by the last lemma. ■

Remark 38 *(a) Given the inequalities $G(c) \leq 0$ and $\lambda \geq 0$, the restriction $\lambda \cdot G(c) = 0$ is known as* ***complimentary slackness****, since it is equivalent to $G(c)_i < 0 \implies \lambda_i = 0$, for every coordinate i. Intuitively, a constraint can have a positive price only if it is binding.*
(b) Condition 2 of Theorem 37 is sometimes equivalently stated as a ***saddle-point condition:*** *$\mathcal{L}(\cdot, \lambda)$ is maximized over C at c, and $\mathcal{L}(c, \cdot)$ is minimized over $\mathbb{R}^n_+$ at λ.*

Assuming the existence of a maximum, the above global optimality conditions are applicable if and only if $\partial J(0)$ is nonempty. Convexity-based sufficient conditions for this to be true are given below. Besides convexity, the key assumption is the so-called **Slater condition:**

$$\text{There exists } x \in C \text{ such that } G(x)_i < 0 \text{ for all } i \in \{1, \ldots, n\}. \tag{12}$$

Proposition 39 *Suppose that $C, -U$ and G are convex, $J(0)$ is finite and the Slater condition (12) holds. Then $\partial J(0) \neq \emptyset$.*

Proof. One can easily verify that sub(J) is convex and has $(0, J(0))$ on its boundary. By the supporting hyperplane theorem, sub(J) is supported at $(0, J(0))$ by some nonzero $(\lambda, -\alpha)$, and therefore $v \leq J(\delta)$ implies $-\alpha J(0) \leq \lambda \cdot \delta - \alpha v$ for all $\delta \in \mathbb{R}^n$. If $\alpha < 0$, then one obtains a contradiction with $\delta = 0$. The Slater condition guarantees that $J(\delta) > -\infty$ for all δ sufficiently close to zero, and therefore $\alpha \neq 0$. The only possibility is $\alpha > 0$, in which case $\lambda / \alpha \in \partial J(0)$. ■

12 Local Optimality Conditions

In this section we develop local optimality conditions in terms of gradients, exploiting the fact that small feasible perturbations near an optimum cannot improve the optimization's objective.

The following local optimality conditions under linear constraints are sufficient for most of this text's applications.

Proposition 40 *Consider a function $U : D \to \mathbb{R}$, where $D \subseteq X$, the set $\{B_1, \ldots, B_n\}$ of linearly independent vectors in X and the linear manifold*

$$M = \{x \in X : (B_i \mid x) = b_i \text{ for } i = 1, \ldots, n\}, \quad \text{where } b \in \mathbb{R}^n.$$

Suppose that the gradient of U at $c \in D^0 \cap M$ exists. Then

$$U(c) = \max_{x \in D \cap M} U(x) \quad \textit{implies} \quad \nabla U(c) = \sum_{i=1}^{n} \lambda_i B_i \textit{ for some } \lambda \in \mathbb{R}^n.$$

The converse implication is also valid if U is assumed to be concave.

Proof. Suppose $U(c) = \max\{U(x) : x \in D \cap M\}$ and consider any vector y in the linear subspace

$$L = M - c = \{x \in X : (B_i \mid x) = 0,\ i = 1, \ldots, n\}.$$

Setting to zero the derivative at zero of $U(c + \alpha y)$ as a function of α results in the orthogonality condition $(\nabla U(c) \mid y) = 0$. Therefore, $\nabla U(c) \in L^{\perp}$. By Proposition 22, $L^{\perp} = M^{\perp} =$ span$(B_1, \ldots, B_n)$ and therefore $\nabla U(c)$ is a linear combination of the B_i. Conversely, suppose U is concave and $\nabla U(c) \in$ span$(B_1, \ldots, B_n)$. Consider any $x \in D \cap M$, and let $y = x - c \in L$. Using the gradient inequality and the fact that $\nabla U(c)$ is orthogonal to y, we conclude that $U(x) \leq U(c) + (\nabla U(c) \mid y) = U(c)$. ■

Finally, we derive the **Kuhn-Tucker optimality conditions** for last section's problem (P_0).

Theorem 41 *Suppose that $c \in C^0$ solves problem $(P_0) : U(c) = J(0)$ and $G(c) \leq 0$. Suppose also that the gradients of U and G at c exist, and that*

$$\textit{for some } h \in X,\ G_i(c) + (\nabla G_i(c) \mid h) < 0 \textit{ for all } i. \tag{13}$$

Then there exists some $\lambda \in \mathbb{R}^n$ such that

$$\nabla U(c) = \lambda \cdot \nabla G(c), \quad \lambda \cdot G(c) = 0, \quad \lambda \geq 0.$$

Proof. An exercise shows that optimality of c implies that for any $h \in X$,

$$G_i(c) + (\nabla G_i(c) \mid h) < 0 \text{ for all } i \quad \Longrightarrow \quad (\nabla U(c) \mid h) \leq 0.$$

The following two convex subsets of $\mathbb{R}^n \times \mathbb{R}$ are therefore nonintersecting:

$$A = \{(\delta, \upsilon) : \delta_i < 0 \text{ for all } i, \text{ and } \upsilon > 0\},$$
$$B = \{(\delta, \upsilon) : \text{for some } h \in X,\ \delta \geq G(c) + (\nabla G(c) \mid h) \text{ and } \upsilon \leq (\nabla U(c) \mid h)\}.$$

By the separating hyperplane theorem (Corollary 33), there exists nonzero $(-\lambda, \alpha) \in \mathbb{R}^n \times \mathbb{R}$ that separates A and B, and therefore

$$\delta \leq 0 \text{ and } \upsilon \geq 0 \quad \Longrightarrow \quad -\lambda \cdot \delta + \alpha \upsilon \geq 0, \tag{14}$$

$$-\lambda \cdot [G(c) + (\nabla G(c) \mid h)] + \alpha (\nabla U(c) \mid h) \leq 0 \quad \text{for all } h \in X. \tag{15}$$

Condition (14) implies that $\alpha \geq 0$ and $\lambda \geq 0$. If $\alpha = 0$, assumption (13) is violated (why?). After proper scaling, we can therefore assume that $\alpha = 1$, and the Kuhn-Tucker conditions follow from condition (15) and the inequalities $\lambda \geq 0$ and $G(c) \leq 0$. ■

13 Exercises

1. Prove Proposition 3.

2. (a) Given the inner product space X, prove the **parallelogram identity**:

$$\|x+y\|^2 + \|x-y\|^2 = 2\|x\|^2 + 2\|y\|^2, \quad x, y \in X.$$

   Interpret this equation geometrically to justify its name.

   (b) Verify the equations in the proof of Theorem 19.

3. Suppose D is an open convex subset of X, $f : D \to \mathbb{R}$ is concave, and $x \in D$. Assume that f is **bounded below near** x, meaning that there exist an open ball $B(x; r) \subseteq D$ and some $b \in \mathbb{R}$ such that $f(y) > b$ for all $y \in B(x; r)$. Fix any $\tilde{x} \in D$, and show that f is bounded below near $\tilde{x}$ by explaining the validity of the following steps:

   (a) There exists some $z \in D$ and $\beta \in (0, 1)$ such that $\tilde{x} = \beta x + (1 - \beta) z$.

   (b) $B(\tilde{x}; \beta r) = \{\beta y + (1 - \beta) z : y \in B(x; r)\}$.

   (c) $B(\tilde{x}; \beta r) \subseteq D$.

   (d) f is bounded below by $\beta b + (1 - \beta) f(z)$ on $B(\tilde{x}; \beta r)$.

4. Let $X = C[0, 1]$ be the vector space of all continuous functions on the unit interval, with the inner product $(x \mid y) = \int_0^1 x(t) y(t)\, dt$.

   (a) Show that the sequence $\{x_n\}$ defined by $x_n(t) = 1/(1 + nt)$, $t \in [0, 1]$, is Cauchy but does not converge in X.

(b) Suppose we are interested in finding a continuous function $x : [0,1] \to \mathbb{R}$ that minimizes $\int_0^1 x(t)^2\, dt$ subject to the constraint $x(0) = 1$. Express this as a projection problem in X, and show that the suitable projection does not exist.

5. Let X be the vector space of every real-valued sequence $x = (x_1, x_2, \ldots)$ such that $x_n = 0$ for all but finitely many values of n, with the inner product $(x \mid y) = \sum_n x_n y_n$. Show that the functional $f(x) = \sum_n n x_n$ is linear but not continuous.

6. Give an alternative proof of the necessity of the support condition in Theorem 19(a) that utilizes the fact that the quadratic $f(\alpha) = \|x - x_\alpha\|^2$ is minimized at zero and therefore the right derivative of f at zero must be nonnegative. Also draw a diagram that makes it obvious that if the angle between $s - x_S$ and $x_S - x$ is wider than a right angle, then there exists some $\alpha \in (0,1)$ such that $\|x - x_\alpha\| < \|x - x_S\|$.

7. (a) Show that the vector y supports the set S at some vector $\bar{s}$ if and only if y supports $\bar{S}$ at $\bar{s}$. (Note that it is not assumed that $\bar{s}$ is in S, and therefore Definition 31 is required.)

   (b) Show that the closure of a convex set is convex and conclude that, in a complete space, projections on the closure of a convex set always exist.

8. Suppose L is a linear subspace of the finite-dimensional vector space X. Show that there exists a basis $\{B_1, \ldots, B_d\}$ of X such that $L = \operatorname{span}(B_1, \ldots, B_k)$ for some $k \leq d$. Using this fact, show that every linear manifold in X is closed, thus proving the existence of projections on linear manifolds in finite dimensions.

9. (a) State the projection expression of Proposition 22 in common matrix notation, assuming that the underlying space is $X = \mathbb{R}^d$ with the Euclidean inner product $(x \mid y) = \sum_{i=1}^n x_i y_i$.

   (b) Prove Corollaries 23 and 24.

10. Prove Proposition 26, and derive the projection expression of Proposition 22 as a corollary of Proposition 9.

11. Suppose P is a linear operator on the inner product space $(X, (\cdot \mid \cdot))$, that is, a function of the form $P : X \to X$ such that $P(\alpha x + \beta y) = \alpha P(x) + \beta P(y)$ for all $\alpha, \beta \in \mathbb{R}$ and $x, y \in X$. P is a **projection operator** if there exists a linear subspace L such that Px is the projection of x on L for all $x \in X$. The operator P

is **idempotent** if $P^2 = P$. Finally, P is **self-adjoint** if $(Px \mid y) = (x \mid Py)$ for all $x, y \in X$.

(a) Prove that P is a projection operator if and only if P is both idempotent and self-adjoint.

(b) Apply part (a) to show that a matrix $A \in \mathbb{R}^{n \times n}$ is idempotent $\left(A^2 = A\right)$ and symmetric $(A' = A)$ if and only if there exists a full-rank matrix B such that $A = B' (BB')^{-1} B$.

12. (a) Suppose X, Y and Z are vector spaces, and $f : X \to Y$ and $g : X \to Z$ are linear functions such that $g(x) = 0$ implies $f(x) = 0$ for all $x \in X$. Show that $L = \{g(x) : x \in X\}$ is a linear subspace of Z, and that there exists a linear function $h : L \to Y$ such that $f = h \circ g$.

    (b) Suppose $f, g_1, \ldots, g_n$ are linear functionals on X such that $g_1(x) = \cdots = g_n(x) = 0$ implies $f(x) = 0$, for all $x \in X$. Show that $f = \sum_{i=1}^{n} \lambda_i g_i$ for some $\lambda \in \mathbb{R}^n$.

13. Use a projection argument to solve the problem

$$\min \left\{x'Qx : Ax = b,\ x \in \mathbb{R}^{n \times 1}\right\},$$

    where $Q \in \mathbb{R}^{n \times n}$ is positive definite and symmetric, $A \in \mathbb{R}^{m \times n}$ is full rank and $b \in \mathbb{R}^m$ is in the range of A.

14. Consider the problem of finding a vector x of minimum norm satisfying $(x \mid y_i) \geq \alpha_i$, $i = 1, \ldots, n$, where the vectors $y_1, \ldots, y_n$ are linearly independent. Use projection arguments to show that the problem has a unique solution and to characterize the solution as a system of (in)equalities.

15. Suppose that the random vector $y = (y_1, \ldots, y_n)'$ is generated by the model $y = A\beta + \varepsilon$, where $A \in \mathbb{R}^{n \times m}$ is a known matrix, $\beta \in \mathbb{R}^m$ is an unknown vector, and ε is an unobservable zero-mean random vector, valued in $\mathbb{R}^n$, with variance-covariance matrix $\mathbb{E}[\varepsilon\varepsilon'] = \Sigma$, assumed to be positive definite. We are interested in a **linear estimator** of the parameter β, that is, an estimator of the form $\hat{\beta} = By$, for some $B \in \mathbb{R}^{m \times n}$. The linear estimator represented by B is **unbiased** if $BA = I$, since then $\mathbb{E}\hat{\beta} = \beta$ for every choice of β. Using projection theory, determine the unbiased linear estimator that minimizes the variance of $\hat{\beta}_i - \beta_i$ for every $i \in \{1, \ldots, m\}$.

16. (∞) Suppose $X = l_2$, the space of square summable sequences with the inner product $(x \mid y) = \sum_{n=1}^{\infty} x(n)\, y(n)$. Let S be the positive cone of X, that is, the set of all $x \in X$ such that $x(n) \geq 0$ for all n. Show that $S = \bar{S}$ and therefore S contains no interior points. Finally, consider any $\bar{s} \in S$ such that $\bar{s}(n) > 0$ for all n, and show that the only vector y such that $(y \mid s) \geq (y \mid \bar{s})$ for all $s \in S$ is the zero vector.

17. Verify the claim of display (11), stating the equivalence of the supergradient property and an associated support condition.

18. Provide a geometric explanation of complementary slackness (defined in Remark 38), based on a support condition relative to the set sub(J) in $\mathbb{R}^n \times \mathbb{R}$.

19. In the context of Section 11, suppose that U and G are continuous. Is J necessarily continuous at an interior point of its effective domain (that is, the set where J is finite)? Provide a proof or a counterexample.

20. Suppose that $f : X \to \mathbb{R}$ is concave.

    (a) Show that $f'(x; h_1 + h_2) \geq f'(x; h_1) + f'(x; h_2)$ for every $h_1, h_2 \in X$ for which the above directional derivatives exist.

    (b) Show that if $f'(x; h)$ exists for all $h \in X$, then the function $f'(x; \cdot)$ is linear.

21. Prove the opening claim of the proof of Theorem 41. Also, provide a set of convexity-based sufficient conditions for the Kuhn-Tucker conditions of Theorem 41 to imply optimality, and prove your claim.

14 Notes

Treatments of finite-dimensional convex optimization theory include Rockafellar (1970), Bertsekas (2003) and Boyd and Vandenberghe (2004), the latter providing an introduction to modern computational optimization algorithms. Infinite-dimensional extensions can be found in Luenberger (1969), Ekeland and Témam (1999) and Bonnans and Shapiro (2000). The Hilbert-space theory outlined in this Appendix is presented and extended in the classic text of Dunford and Schwartz (1988), where one can also find the notion of weak compactness and its interplay with convexity alluded to in Section 8.

References

Bertsekas, D. P. (2003): *Convex Analysis and Optimization*. Athena Scientific, Belmont, MA.

Bonnans, J. F., and A. Shapiro (2000): *Perturbation Analysis of Optimization Problems*. Springer-Verlag, New York.

Boyd, S., and L. Vandenberghe (2004): *Convex Optimization*. Cambridge University Press, Cambridge, United Kingdom.

Dunford, N., and J. T. Schwartz (1988): *Linear Operators, Part I, General Theory*. Wiley.

Ekeland, I., and R. Témam (1999): *Convex Analysis and Variational Problems*. SIAM, Philadelphia, PA.

Luenberger, D. G. (1969): *Optimization by Vector Space Methods*. Wiley, New York.

Rockafellar, R. T. (1970): *Convex Analysis*. Princeton University Press, Princeton, New Jersey.

Skiadas, C. (2009): *Asset Pricing Theory*. Princeton Univ. Press, Princeton, NJ.