2.1 Introduction

This chapter is devoted to characterizing best approximations in normed spaces. In a normed space, generally it is through the norm of the space that best approximations are characterized. Here we take a different approach. Instead of the norm, it is a semi-inner product which generates the norm of the space, that is employed in characterizing best approximations. This method enables us to derive some characterizations of best approximations in normed spaces which are entirely new.

We begin this chapter with a brief discussion on best approximation in normed spaces. In our attempt to characterize best approximations, we first of all derive a result characterizing best approximations from convex sets using the defining properties of semi-inner products and convexity of sets. The notions of dual cone and orthogonal complement of a set are then introduced, and some of their basic properties are discussed. From a reformulated version of the characterization result for convex sets in terms of dual cones, characterizations of best approximations from convex cones and subspaces are arrived at. Characterizing best approximations from translates of convex cones and subspaces are also
2.2 Best Approximation in Normed Spaces

It is well known that the problem of best approximation of a function consists of the determination of a function belonging to a fixed family such that its deviation from the given function is a minimum. This problem was first formulated in 1853 by P. L. Chebyshev, who investigated the approximation of continuous functions by algebraic polynomials of given degree. As a measure of the deviation between two functions, he used the maximum of the absolute values of their difference. Subsequently, a number of mathematicians have started studying other specialized problems of best approximation. With the development of the theory of normed spaces, it became clear that a wide range of problems of best approximation can be put into a general formulation in terms of normed spaces, if the norm of the space is taken as the measure of deviation. This formulation made possible the application of the methods and ideas of functional analysis and geometry to the problems of approximation theory.

The foundations of the theory of best approximation in normed spaces were established in 1920’s by one of the founders of functional analysis, S. Banach. During 1930-1950, the ideas of Banach were developed and systematized by the mathematicians like S. M. Nicolescu, M. G. Krein, N. I. Achiezer, A. I. Markushhevich, J. L. Walsh and A. N. Kolmogorov.

The problem of best approximation in a normed space can be formulated as follows: Let \((X, \|\cdot\|)\) be a normed space over the real or complex number field \(\mathbb{K}\), \(G\) a nonempty set in \(X\), and \(x \in X\). Then the distance of \(x\) from \(G\), \(d(x, G)\), is given by

\[
d(x, G) := \inf \{\|x - g\| : g \in G\}.
\]

The problem of best approximation consists of finding an element \(g_0 \in G\) such that

\[
\|x - g_0\| = d(x, G).
\]

Every element \(g_0 \in G\) with this property is called a best approximation of \(x\) from
2.2 Best Approximation in Normed Spaces

Let \((X, \|\cdot\|)\) be a normed space over the real or complex number field \(\mathbb{K}\), \(G\) a nonempty subset of \(X\), and \(x \in X\). An element \(g_0 \in G\) is called a best approximation (or element of best approximation or nearest point) of \(x\) from \(G\), if \(\|x - g_0\| = d(x, G)\). In this case, the number \(d(x, G)\) is called the error of approximation (or the error in approximating \(x\) by \(G\)).

An element \(g_0 \in G\) is a best approximation of \(x\) from \(G\) if and only if \(\|x - g_0\| \leq \|x - g\|\) for every \(g \in G\). The set of all best approximations of \(x\) from \(G\) is denoted by \(P_G(x)\) Thus

\[
P_G(x) := \{g_0 \in G : \|x - g_0\| = d(x, G)\}.
\]

This defines a mapping \(P_G : X \to \mathcal{P}(G)\), where \(\mathcal{P}(G)\) is the power set of \(G\). The set valued mapping \(P_G\) is called the metric projection (or nearest point mapping or proximity map) onto \(G\).

If each \(x \in X\) has at least (respectively at most) one best approximation from \(G\), then \(G\) is called a proximinal (respectively semi Chebyshev) set. If each \(x \in X\) has exactly one best approximation from \(G\), then \(G\) is called a Chebyshev set. Thus, \(G\) is proximinal (respectively semi Chebyshev, Chebyshev) if \(P_G(x)\) is nonempty (respectively \(P_G(x)\) is either empty or a singleton, \(P_G(x)\) is a singleton) for each \(x \in X\). It is obvious that a Chebyshev set is proximinal as well as semi Chebyshev. If \(G\) is a Chebyshev set in \(X\), then the metric projection \(P_G\) is a single valued mapping of \(X\) onto \(G\), and in this case, \(P_G\) is called the Chebyshev map (or best approximation operator) onto \(G\).

The general theory of best approximation may be briefly outlined as follows: It is the mathematical study that is motivated by the desire to seek answers to the following basic questions, among others.

1. (Existence of best approximations) Which subsets are proximinal?
2. (Uniqueness of best approximations) Which subsets are Chebyshev?
3. (Characterization of best approximations) How does one recognize when a given element \(g \in G\) is a best approximation of \(x\) from \(G\)?
4. (Error of approximation) How does one compute the error of approximation \( d(x, G) \), or at least get sharp upper or lower bounds for it?

5. (Computation of best approximations) Can one describe some useful algorithms for actually computing best approximations?

6. (Continuity of best approximations) How does \( P_G(x) \) vary as a function of \( x \) (or \( G \))?

Many mathematicians were attracted by these questions and have made their contributions to the theory of best approximation. As a result, a pretty large collection of materials including Textbooks, Treaties, Monographs and Papers is there in the literature (e.g., A. L. Brown [6], P. L. Butzer and R. J. Nessel [7], W. Cheney and W. Light [10], E. W. Cheney and K. H. Price [8,9], E. W. Cheney and P. D. Morris [11], F. Deutsch [14], R. A. DeVore [32], R. A. DeVore and G. G. Lorentz [33], W. O. G. Lewicki [16], H. N. Mhaskar and D. V. Pai [21], M. J. D. Powell [24], T. J. Rivlin [26], H. S. Shapiro [28], I. Singer [29,30], G. A. Watson [34], R. Zielke [35] and so on).

Among the six questions mentioned above, the question which is of particular interest to us is the third one, that is, the characterization of best approximations. We understand that the literature is rich with results characterizing best approximations, and that such results are separately available for general normed spaces and inner product spaces (e.g., H. N. Mhaskar and D. V. Pai [21], I. Singer [29,30], H. Berens [4], F. Deutsch [14] and so on). Generally, characterizations of best approximations in a normed space are derived through the norm of the space, and those in an inner product space, through the inner product of the space. Here we take a different approach. Our endeavor is to characterize best approximations in a general normed space, not through the norm of the space, but through a semi-inner product that generates the norm of the space. Our attention here is to derive some results characterizing best approximations in the framework of a general normed space using some semi-inner product techniques.

We recall from Chapter 1 that given a normed space \((X, \|\cdot\|)\) over the real or complex number field \(\mathbb{K}\), there always exists a semi-inner product (see Defi-
nition 1.4.3) on it which generates the norm \( \| \cdot \| \). This idea is employed here in characterizing best approximations in normed spaces from convex sets, and in particular, from convex cones and subspaces.

Since, for every nonempty subset \( G \) of a normed space \( X \), we have

\[
P_G(x) = \begin{cases} 
  x & \text{if } x \in G \\
  \emptyset & \text{if } x \in \overline{G} \setminus G,
\end{cases}
\]

it is sufficient to characterize the best approximations of the elements \( x \in X \setminus \overline{G} \). In order to exclude the trivial case when such elements \( x \) do not exist, throughout the discussion, our approximating sets \( G \), whether convex sets, convex cones or subspaces, as the case may be, are always assumed to be proper and nondense in the normed space \( X \). However, we will not be making any special mention to these effects in the sequel.

Our further discussion is restricted to the setting of real normed spaces. Henceforth in this chapter, by a normed space \( X \) we mean a real normed space \((X, \| \cdot \|)\) together with a semi-inner product \([\cdot | \cdot]\) which generates the norm \( \| \cdot \| \).

Since the theory of best approximation is the most well developed when the approximating set is a subspace, or more generally a convex set, we begin our study with the characterization of best approximations from convex sets.

### 2.3 Characterization from Convex Sets

The sole aim of this section is to present a characterization theorem for best approximations from convex sets, and to reformulate it in terms of dual cones of sets. This result will prove useful over and over again throughout our discussion. Indeed, it will be the basis for every characterization theorem that we provide.

We begin our discussion with a sufficient condition for best approximations from arbitrary sets. Recall that \( P_G(x) \) denotes the set of all best approximations of \( x \) from \( G \).
2.3 Characterization from Convex Sets

Theorem 2.3.1. Let $X$ be a normed space, $G$ a subset of $X$, $x \in X$, and $y_0 \in G$. If

\[(2.1) \quad [y - y_0|x - y_0] \leq 0 \quad \text{for all } y \in G,\]

then $y_0 \in P_G(x)$.

Proof. Suppose that (2.1) holds. Then for all $y \in G$, we have

\[\|x - y_0\|^2 = [x - y_0|x - y_0], \quad \text{by } (S_3)\]
\[= [x - y|x - y] + [y - y_0|x - y_0], \quad \text{by } (S_1)\]
\[\leq [x - y|x - y], \quad \text{by } (2.1)\]
\[\leq \|x - y\|\|x - y_0\|, \quad \text{by } (S_4).\]

Hence $\|x - y_0\| \leq \|x - y\|$ for all $y \in G$, and so $y_0 \in P_G(x)$. \qed

When the approximating set is in particular a convex set, we have the following necessary condition for best approximations.

Theorem 2.3.2. Let $X$ be a normed space, $K$ a convex subset of $X$, $x \in X$, and $y_0 \in K$. If $y_0 \in P_K(x)$, then

\[(2.2) \quad [y - y_0|x - y_0 - \lambda(y - y_0)] \leq 0 \quad \text{for all } y \in K \text{ and all } \lambda \in [0, 1].\]

Proof. If (2.2) fails, then for some $y \in K$ and some $\lambda \in [0, 1]$, we have,

\[(2.3) \quad [y - y_0|x - y_0 - \lambda(y - y_0)] > 0.\]

(Here $y \neq y_0$, since $y = y_0$ implies $[y - y_0|x - y_0 - \lambda(y - y_0)] = 0$.) For these elements $y \in K$ and $\lambda \in [0, 1]$, the element $y_\lambda := \lambda y + (1 - \lambda) y_0 \in K$, by convexity of $K$. We have then

\[\|x - y_\lambda\| = \|x - \lambda y - (1 - \lambda) y_0\|\]
\[= \frac{[x - y_0 - \lambda(y - y_0)|x - y_0 - \lambda(y - y_0)]}{\|x - y_0 - \lambda(y - y_0)\|}, \quad \text{by } (S_3)\]
2.3 Characterization from Convex Sets

\[
\frac{[x - y_0]|x - y_0 - \lambda (y - y_0)|}{\|x - y_0 - \lambda (y - y_0)\|}, \text{ by } (S_1)
\]

\[
< \frac{[x - y_0]|x - y_0 - \lambda (y - y_0)|}{\|x - y_0 - \lambda (y - y_0)\|}, \text{ by } (2.3)
\]

\[
\leq \|x - y_0\|, \text{ by } (S_4).
\]

This shows that there is a \( y_\lambda \in K \) such that \( \|x - y_\lambda\| < \|x - y_0\| \), and so \( y_0 \notin P_K(x) \). Hence (2.2) holds whenever \( y_0 \in P_K(x) \). \( \square \)

Combining Theorem 2.3.1 and Theorem 2.3.2, we have the following result which characterizes best approximations from convex sets.

**Theorem 2.3.3.** Let \( X \) be a normed space, \( K \) a convex subset of \( X \), \( x \in X \), and \( y_0 \in K \). Then the following statements are equivalent:

(a) \( y_0 \in P_K(x); \)

(b) \( [(y - y_0) \mid x - y_0 - \lambda (y - y_0)] \leq 0 \) for all \( y \in K \) and all \( \lambda \in [0, 1]; \)

(c) \( [y - y_0 \mid x - y_0] \leq 0 \) for all \( y \in K. \)

**Proof.** (a) \( \Rightarrow \) (b) follows by Theorem 2.3.2, (b) \( \Rightarrow \) (c) follows by taking \( \lambda = 0 \) in (b), and (c) \( \Rightarrow \) (a) follows by Theorem 2.3.1. \( \square \)

There is yet another way of stating the above characterization result. It involves the notion of the dual cone of a given set. This notion has been introduced in the framework of an inner product space in terms of the inner product of the space [14]. We extend this to the setting of a normed space in terms of a semi-inner product that generates the norm of the space.

**Definition 2.3.4.** Let \( X \) be a normed space, and \( G \) a nonempty subset of \( X \). Then the set \( \{ x \in X : [y|x] \leq 0 \text{ for all } y \in G \} \) is called the dual cone (or dual cone relative to the semi-inner product \([\cdot|\cdot]\), or negative polar relative to the semi-inner product \([\cdot|\cdot]\)) of \( G \), denoted by \( G^\circ \).

By the definition, for every nonempty subset \( G \) of \( X \), \( 0 \in G^\circ \) and \( G \cap G^\circ \) is either empty or \( \{0\} \).
Theorem 2.3.3 characterizing best approximations from convex sets can now be reformulated using dual cones as follows.

**Theorem 2.3.5.** Let \( X \) be a normed space, \( K \) a convex subset of \( X \), \( x \in X \), and \( y_0 \in K \). Then the following statements are equivalent:

(i) \( y_0 \in P_K(x) \);

(ii) \( x - y_0 - \lambda(y - y_0) \in (K - y_0)^\circ \) for all \( y \in K \) and all \( \lambda \in [0, 1] \);

(iii) \( x - y_0 \in (K - y_0)^\circ \).

**Proof.** From the equivalence of statements (a) and (b) of Theorem 2.3.3 and by the definition of dual cone of a set (Definition 2.3.4), we have

\[
y_0 \in P_K(x) \iff [y - y_0|x - y_0 - \lambda(y - y_0)|] \leq 0 \text{ for all } y \in K \text{ and all } \lambda \in [0, 1]
\]

\[
\iff x - y_0 - \lambda(y - y_0) \in \{y - y_0 : y \in K\}^\circ \text{ for all } y \in K \text{ and all } \lambda \in [0, 1]
\]

\[
\iff x - y_0 - \lambda(y - y_0) \in (K - y_0)^\circ \text{ for all } y \in K \text{ and all } \lambda \in [0, 1].
\]

Hence (i) \(\iff\) (ii).

Similarly, from the equivalence of statements (a) and (c) of Theorem 2.3.3, and by the definition of dual cone of a set, we have

\[
y_0 \in P_K(x) \iff [y - y_0|x - y_0|] \leq 0 \text{ for all } y \in K
\]

\[
\iff x - y_0 \in \{y - y_0 : y \in K\}^\circ
\]

\[
\iff x - y_0 \in (K - y_0)^\circ.
\]

Hence (i) \(\iff\) (iii), and this completes the proof.

The above theorem shows that the characterization of best approximations requires, in essence, the determination of dual cones. For certain convex sets (e.g., convex cones and subspaces), substantial improvements in Theorem 2.3.5 are possible.
Definition 2.3.6. Let $X$ be a normed space. A subset $G$ of $X$ is called a convex cone if $\alpha x + \beta y \in G$ whenever $x, y \in G$ and $\alpha, \beta \geq 0$.

We make the following observations:

1. If $C$ is a convex cone, then $0 \in C$ and $C \cap C^o = \{0\}$.

2. Every subspace is a convex cone, but not conversely. Similarly, every convex cone is a convex set, but not conversely. For instance, in the real normed space $X = (\mathbb{R}^2, \|\cdot\|_2)$, where $\|x\|_2 = \left(\sum_{k=1}^{2} |x_k|^2\right)^{1/2}$ for $x = (x_1, x_2) \in X$, the set $\{x = (x_1, x_2) \in X : x_1 \geq 0, x_2 \geq 0\}$ is a convex cone which is not a subspace, and the closed unit ball $\{x \in X : \|x\|_2 \leq 1\}$ is a convex set which is not a convex cone.

3. Generally, the dual cone of a nonempty set is not a convex cone, and the same is the case with that of a convex cone also. For example, consider the real normed space $X = (\mathbb{R}^2, \|\cdot\|_1)$, where $\|x\|_1 = \sum_{k=1}^{2} |x_k|$ for $x = (x_1, x_2) \in X$, together with the semi-inner product

$$[x|y] = \|y\|_1 \sum_{k=1}^{2} \frac{x_k y_k}{|y_k|}, \quad x, y \in X$$

that generates the norm $\|\cdot\|_1$. The set

$$C = \{x = (\lambda, \lambda) \in X : \lambda \geq 0\}$$

is a convex cone in $X$ whose dual cone is given by

$$C^o = \left\{ x = (x_1, x_2) \in X : \lambda \|x\|_1 \sum_{k=1}^{2} \frac{x_k |x_k|}{|x_k|} \leq 0 \text{ for all } \lambda \geq 0 \right\}.$$ 

Then the elements $x = (5, -4)$ and $z = (-3, 5)$ are in $C^o$. But $x + z = (2, 1) \notin C^o$. This shows that $C^o$ is not a convex cone.
2.4 Orthogonality relative to Semi-Inner Products

In this section we introduce the notion of orthogonality in a normed space in terms of a semi-inner product that generates the norm of the space [15]. This will enable us to derive some basic properties of dual cones that are needed for improving Theorem 2.3.5.

The notion of orthogonality in a normed space which we discuss here is a generalization of the orthogonality concept in an inner product space. Recall that two elements $x, y$ in an inner product space $(X, (\cdot, \cdot))$ are said to be orthogonal if $(x, y) = 0$. Since a semi-inner product lacks conjugate symmetry, a property which an inner product possesses, and since it is in terms of a semi-inner product which generates the norm of the space that we introduce orthogonality here, this notion of orthogonality in a normed space is not generally symmetric. Unless specified otherwise, by orthogonality in a normed space we always mean this orthogonality, as defined below.

**Definition 2.4.1.** Let $X$ be a normed space, and $x, y \in X$. Then $x$ is said to be orthogonal (or orthogonal in the sense of Lumer-Giles relative to the semi-inner product $[-,-]$) to $y$, denoted by $x \perp y$, if $[y|x]=0$.

We observe that, if $x, y, z \in X$, and $\alpha$ is any scalar, then

(i) $0 \perp x$ and $x \perp 0$,
(ii) $x \perp x$ if and only if $x = 0$,
(iii) $x \perp y$ and $x \perp z$ imply that $x \perp (y + z)$, and
(iv) $x \perp y$ implies that $(\alpha x) \perp y$ and $x \perp (\alpha y)$.

However, as we have mentioned above, the main difference of this orthogonality concept in normed spaces in comparison with the orthogonality concept in inner product spaces is with regard to symmetry: If $x, y \in X$, then $x \perp y$ need not imply that $y \perp x$.

For example, consider the real normed space $X = (\mathbb{R}^3, \|\cdot\|_1)$, where $\|x\|_1 =$
$\sum_{k=1}^{3} |x_k|$ for $x = (x_1, x_2, x_3) \in X$, equipped with the semi-inner product

$$[x|y] = \|y\|_1 \sum_{k=1}^{3} \frac{x_k y_k}{|y_k|}, \quad x, y \in X$$

that generates the norm $\|\cdot\|_1$. Consider the elements $x = (-2, 1, 0)$ and $y = (1, 1, 0)$ in $X$. Then $[y|x] = 0$ so that $x$ is orthogonal to $y$, whereas $[x|y] = -2$ so that $y$ is not orthogonal to $x$.

**Definition 2.4.2.** Let $X$ be a normed space, $G$ a nonempty subset of $X$, and $x \in X$. Then $x$ is said to be orthogonal (or orthogonal in the sense of Lumer-Giles relative to the semi-inner product $[\cdot|\cdot]$) to $G$, denoted by $x \perp G$, if $x \perp y$ for all $y \in G$.

For every nonempty subset $G$ of $X$, by the definition, $0 \perp G$.

**Definitions 2.4.3.** Let $X$ be a normed space, and $G$ a nonempty subset of $X$. Then the set $\{x \in X : x \perp G\}$ is called the orthogonal complement (or orthogonal complement in the sense of Lumer-Giles relative to the semi-inner product $[\cdot|\cdot]$) of $G$, denoted by $G^\perp$.

If $y \in X$, the orthogonal complement (or orthogonal complement in the sense of Lumer-Giles relative to the semi-inner product $[\cdot|\cdot]$) of $y$, denoted by $y^\perp$, is the set $\{x \in X : x \perp y\}$.

We have

$$G^\perp = \{x \in X : x \perp G\}$$

$$= \{x \in X : x \perp y \text{ for all } y \in G\}$$

$$= \bigcap_{y \in G} \{x \in X : x \perp y\}$$

$$= \bigcap_{y \in G} y^\perp.$$

The following are some easy consequences of these definitions:

1. $0^\perp = X$, and $X^\perp = \{0\}$. 
2. If $G$ is any nonempty subset of $X$, and $\alpha$ is any scalar, then,

(a) $0 \in G^\perp$,
(b) $x \in G^\perp$ implies that $\alpha x \in G^\perp$,
(c) $G^\perp \subseteq G^\circ$, the dual cone of $G$, and
(d) $G \cap G^\perp$ is either empty or $\{0\}$.

3. If $C$ is a convex cone in $X$, we also have $C \cap C^\perp = \{0\}$. In particular, $M \cap M^\perp = \{0\}$ for every subspace $M$ of $X$.

4. More importantly, even if $M$ is a subspace of $X$, $M^\perp$ need not be a subspace of $X$. $M^\perp$ is not even a convex cone in $X$. For example, consider the real normed space $X = (R^2, \|\cdot\|_1)$, where $\|x\|_1 = \sum_{k=1}^{2} |x_k|$ for $x = (x_1, x_2) \in X$, together with the semi-inner product

$$[x|y] = \|y\|_1 \sum_{k=1}^{2} \frac{x_k y_k}{|y_k|}, \quad x, y \in X$$

that generates the norm $\|\cdot\|_1$. Let $M = \text{span}\{(1,1)\}$. The orthogonal complement of this subspace is given by

$$M^\perp = \left\{ x = (x_1, x_2) \in X : \lambda \|x\|_1 \sum_{k=1}^{2} \frac{x_k}{|x_k|} = 0 \text{ for all } \lambda \in \mathbb{R} \right\}.$$ 

Then the elements $x = (-3, 5)$ and $z = (4, -3)$ are in $M^\perp$. But $x + z = (1, 2) \notin M^\perp$. This shows that $M^\perp$ is not a convex cone.

Among these observations, the one which is mentioned last is the crucial difference in comparison with the usual orthogonal complements in inner product spaces.

The exact relationship between the dual cone and the orthogonal complement of a given set is provided by the following result.

**Theorem 2.4.4.** Let $X$ be a normed space, and $G$ a nonempty subset of $X$. Then $G^\perp = G^\circ \cap (-G)^\circ = G^\circ \cap (-G^\circ)$. 
2.4 Orthogonality relative to Semi-Inner Products

Proof. We have

\[ G^\perp = \{ x \in X : x \perp G \} \]
\[ = \{ x \in X : [y|x] = 0 \text{ for all } y \in G \} \]
\[ = G^o \cap \{ x \in X : [y|x] \geq 0 \text{ for all } y \in G \} \]
\[ = G^o \cap \{ x \in X : [-y|x] \geq 0 \text{ for all } -y \in G \} \]
\[ = G^o \cap \{ x \in X : -[y|x] \geq 0 \text{ for all } y \in (-G) \} \]
\[ = G^o \cap \{ x \in X : [y|x] \leq 0 \text{ for all } y \in (-G) \} \]
\[ = G^o \cap (-G)^o. \]

Further,

\[ (-G)^o = \{ x \in X : [y|x] \leq 0 \text{ for all } y \in (-G) \} \]
\[ = \{ x \in X : [y|x] \leq 0 \text{ for all } -y \in G \} \]
\[ = \{ x \in X : [-y|x] \leq 0 \text{ for all } y \in G \} \]
\[ = \{ x \in X : [y|(-x)] \leq 0 \text{ for all } y \in G \} \]
\[ = \{ x \in X : [y|x] \geq 0 \text{ for all } y \in (-G) \} \]
\[ = -\{ -x \in X : [y|(-x)] \leq 0 \text{ for all } y \in G \} \]
\[ = -G^o. \]

Hence \( G^\perp = G^o \cap (-G)^o = G^o \cap (-G^o). \)

Next we consider some basic properties of dual cones and orthogonal complements. The following result will help us in improving Theorem 2.3.5.

**Theorem 2.4.5.** Let \( X \) be a normed space.

(a) If \( C \) is a convex cone in \( X \), then \( (C - y)^o = C^o \cap y^\perp \) for every \( y \in C \).

(b) If \( M \) is a subspace of \( X \), then \( M^o = M^\perp \).

Proof. (a) Let \( y \in C \) be arbitrary. If \( x \in (C - y)^o \), then \([c - y|x] \leq 0 \) for all \( c \in C \). This implies, on letting \( c = c + y \) that \([c|x] = [c + y - y|x] \leq 0 \) for all \( c \in C \), so that \( x \in C^o \). Again, \([c - y|x] \leq 0 \) for all \( c \in C \) implies, on taking \( c = 2y \) and \( c = 0 \), that \([y|x] \leq 0 \) and \([y|x] \geq 0 \) respectively. Thus \([y|x] = 0 \), so that \( x \in y^\perp \). Combining the above two conclusions we see that, \( x \in (C - y)^o \) implies \( x \in C^o \) and \( x \in y^\perp \), so that \( x \in C^o \cap y^\perp \).
On the other hand, if \( x \in C^\circ \cap y^\perp \), then \( c|x| \leq 0 \) for all \( c \in C \) and \( y|x| = 0 \). Consequently, \( [c - y|x|] = c|x| - [y|x|] \leq 0 \) for all \( c \in C \), so that \( x \in (C - y)^\circ \). Hence \( (C - y)^\circ = C^\circ \cap y^\perp \) for every \( y \in C \).

(b) If \( M \) is a subspace of \( X \), then \( -M = M \), and hence by Theorem 2.4.4, \( M^\perp = M^\circ \cap (-M)^\circ = M^\circ \). This completes the proof.

Another result which we will make use of in our further discussion is given below. By \( \overline{G} \) we denote the closure of a set \( G \) under the norm.

**Theorem 2.4.6.** Let \( X \) be a normed space, and \( G \) a nonempty subset of \( X \). Then

(a) \( G^\circ = (\overline{G})^\circ \), and

(b) \( G^\perp = (\overline{G})^\perp \).

**Proof.** (a) Let \( x \in (\overline{G})^\circ \). Then for all \( y \in \overline{G} \), we have \( [y|x] \leq 0 \). This shows, since \( G \subseteq \overline{G} \), that \( [y|x] \leq 0 \) for all \( y \in G \), so that \( x \in G^\circ \). Thus \( (\overline{G})^\circ \subseteq G^\circ \). Now let \( x \in G^\circ \). If \( y \in \overline{G} \), choose a sequence \((y_n)\) in \( G \) such that \( y_n \to y \). Then

\[
|[y_n|x] - [y|x]| = |[y_n - y|x]| \leq \|y_n - y\| \|x\| \to 0 \quad \text{as} \quad n \to \infty,
\]

so that

\( [y|x] = \lim_{n \to \infty} [y_n|x] \leq 0 \).

Therefore \( [y|x] \leq 0 \) for all \( y \in \overline{G} \), so that \( x \in (\overline{G})^\circ \). Thus \( G^\circ \subseteq (\overline{G})^\circ \). Hence \( G^\circ = (\overline{G})^\circ \).

(b) An argument similar to the above shows that \( G^\perp = (\overline{G})^\perp \). \( \square \)

The following result, which is of independent interest, contains some more properties of dual cones and orthogonal complements. We recall that the *sum* of a finite collection of nonempty sets \( \{G_1, G_2, \ldots, G_n\} \) in a normed space \( X \), denoted by \( G_1 + G_2 + \ldots + G_n \) or \( \sum_{i=1}^{n} G_i \), is defined as the set \( \left\{ \sum_{i=1}^{n} x_i : x_i \in G_i \quad \text{for every} \quad i \right\} \). Thus

\[
\sum_{i=1}^{n} G_i := \left\{ \sum_{i=1}^{n} x_i : x_i \in G_i \quad \text{for every} \quad i \right\}.
\]
2.4 Orthogonality relative to Semi-Inner Products

Theorem 2.4.7. Let $X$ be a normed space, and \{$G_1, G_2, ..., G_n$\} a finite collection of nonempty sets in $X$. Then

(a) \( \left( \bigcup_{i=1}^{n} G_i \right)^\circ = \bigcap_{i=1}^{n} G_i^\circ \) and \( \left( \bigcup_{i=1}^{n} G_i \right)^\perp = \bigcap_{i=1}^{n} G_i^\perp \).

(b) If, in addition $0 \in \bigcap_{i=1}^{n} G_i$, then

\( \left( \bigcup_{i=1}^{n} G_i \right)^\circ = \left( \sum_{i=1}^{n} G_i \right)^\circ \) and \( \left( \bigcup_{i=1}^{n} G_i \right)^\perp = \left( \sum_{i=1}^{n} G_i \right)^\perp \).

Proof. (a) We have

\[
x \in \bigcap_{i=1}^{n} G_i^\circ \iff x \in G_i^\circ \text{ for each } i
\]

\[
\iff [y|x] \leq 0 \text{ for each } y \in G_i \text{ and all } i
\]

\[
\iff [y|x] \leq 0 \text{ for all } y \in \bigcup_{i=1}^{n} G_i
\]

\[
\iff x \in \left( \bigcup_{i=1}^{n} G_i \right)^\circ.
\]

Hence \( \left( \bigcup_{i=1}^{n} G_i \right)^\circ = \bigcap_{i=1}^{n} G_i^\circ \).

Similarly, \( \left( \bigcup_{i=1}^{n} G_i \right)^\perp = \bigcap_{i=1}^{n} G_i^\perp \).

(b) We have

\[
x \in \left( \sum_{i=1}^{n} G_i \right)^\circ \iff [y|x] \leq 0 \text{ for each } y \in \sum_{i=1}^{n} G_i
\]

\[
\iff \left[ \sum_{i=1}^{n} y_i |x \right] \leq 0 \text{ whenever } y_i \in G_i
\]

\[
\iff [y_i|x] \leq 0 \text{ for each } y_i \in G_i \text{ and all } i, \text{ since } 0 \in \bigcap_{i=1}^{n} G_i
\]

\[
\iff x \in G_i^\circ \text{ for each } i.
\]
2.5 Characterization from Convex Cones and Subspaces

\[ \Leftrightarrow x \in \bigcap_{i=1}^{n} G_i^\circ \]

\[ \Leftrightarrow x \in \left( \bigcup_{i=1}^{n} G_i \right)^\circ, \text{ by (a) above.} \]

Hence \( \left( \bigcup_{i=1}^{n} G_i \right)^\circ = \left( \sum_{i=1}^{n} G_i \right)^\circ \).

Similarly, \( \left( \bigcup_{i=1}^{n} G_i \right)^\perp = \left( \sum_{i=1}^{n} G_i \right)^\perp \).

\[ \Box \]

\section{2.5 Characterization from Convex Cones and Subspaces}

Using the results of the above section, we now proceed to improve Theorem 2.3.5 characterizing best approximations from convex sets. In the particular case when the convex set is a convex cone, this result can be strengthened by using Theorem 2.4.5 (a) as follows.

\textbf{Theorem 2.5.1.} Let \( X \) be a normed space, \( C \) a convex cone in \( X \), \( x \in X \), and \( y_0 \in C \). Then the following statements are equivalent:

\begin{enumerate}
\item[(a)] \( y_0 \in \mathcal{P}_C(x) \);
\item[(b)] \( [y']|x - y_0 - \lambda(y - y_0)| \leq 0 \) and \( \left[ y_0 |x - y_0 - \lambda(y - y_0) \right] = 0 \) for all \( y' \), \( y \in C \) and all \( \lambda \in [0, 1] \);
\item[(c)] \( [y|x - y_0] \leq 0 \) and \( \left[ y_0 |x - y_0 \right] = 0 \) for all \( y \in C \).
\end{enumerate}

\textbf{Proof.} From the equivalence of statements (i) and (ii) of Theorem 2.3.5, we have

\[ y_0 \in \mathcal{P}_C(x) \Leftrightarrow x - y_0 - \lambda(y - y_0) \in (C - y_0)^\circ \text{ for all } y \in C \text{ and all } \lambda \in [0, 1] \]

\[ \Leftrightarrow x - y_0 - \lambda(y - y_0) \in C^\circ \cap y_0^\perp \text{ for all } y \in C \text{ and all } \lambda \in [0, 1], \text{ by Theorem 2.4.5(a)} \]

\[ \Leftrightarrow x - y_0 - \lambda(y - y_0) \in C^\circ \text{ and } x - y_0 - \lambda(y - y_0) \in y_0^\perp \text{ for all } y \in C \text{ and all } \lambda \in [0, 1] \]
2.5 Characterization from Convex Cones and Subspaces

\[\Leftrightarrow \quad [y' | x - y_0 - \lambda(y - y_0)] \leq 0 \quad \text{and} \quad [y_0 | x - y_0 - \lambda(y - y_0)] = 0 \]
for all \( y', y \in C \) and all \( \lambda \in [0,1] \).

Hence (a)\(\Leftrightarrow\)(b).

Similarly, from the equivalence of statements (i) and (iii) of Theorem 2.3.5, it follows that (a)\(\Leftrightarrow\)(c). Hence the theorem.

\[\text{Remark 2.5.2.} \quad \text{The above theorem can be expressed in a purely set theoretic manner as follows. Under the hypothesis of Theorem 2.5.1, the statements} \]

\( \text{(a) } y_0 \in P_C(x), \)
\( \text{(b) } x - y_0 - \lambda(y - y_0) \in C^\circ \cap y_0^\perp \text{ for all } y \in C \text{ and all } \lambda \in [0,1], \) and
\( \text{(c) } x - y_0 \in C^\circ \cap y_0^\perp. \)

are equivalent.

There is an even simpler characterization of best approximations when the convex set is actually a subspace. It is derived again from Theorem 2.3.5 with the aid of Theorem 2.4.5 (b).

\[\text{Theorem 2.5.3.} \quad \text{Let } X \text{ be a normed space, } M \text{ a subspace of } X, x \in X, \text{ and } y_0 \in M. \text{ Then the following statements are equivalent:} \]

\( \text{(a) } y_0 \in P_M(x), \)
\( \text{(b) } [y' | x - y_0 - \lambda(y - y_0)] = 0 \text{ for all } y', y \in M \text{ and all } \lambda \in [0,1]; \)
\( \text{(c) } [y | x - y_0] = 0 \text{ for all } y \in M. \)

\[\text{Proof. From the equivalence (i)\(\Leftrightarrow\)(ii) of Theorem 2.3.5, we have} \]
\( y_0 \in P_M(x) \Leftrightarrow x - y_0 - \lambda(y - y_0) \in (M - y_0)^\circ \text{ for all } y \in M \text{ and} \)
all \( \lambda \in [0,1] \)
\[\Leftrightarrow x - y_0 - \lambda(y - y_0) \in M^\circ \text{ for all } y \in M \text{ and all } \lambda \in [0,1], \]
since \( y_0 \in M \)
\[\Leftrightarrow x - y_0 - \lambda(y - y_0) \in M^\perp \text{ for all } y \in M \text{ and all } \lambda \in [0,1], \]
by Theorem 2.4.5(b).
\[ \iff \begin{align*} & [y'|x - y_0 - \lambda(y - y_0)] = 0 \text{ for all } y', y \in M \text{ and } \\
& \text{all } \lambda \in [0,1]. \end{align*} \]

Hence (a)⇔(b).

Similarly, from the equivalence of statements (i) and (iii) of Theorem 2.3.5, it follows that (a)⇔(c), and this completes the proof.

**Remark 2.5.4.** The following is the restatement of the above theorem in terms of orthogonal complements. Under the hypothesis of Theorem 2.5.3, the statements

\begin{enumerate}
\item[(a)] \( y_0 \in P_M(x) \),
\item[(b)] \( x - y_0 - \lambda(y - y_0) \in M^\perp \) for all \( y \in M \) and all \( \lambda \in [0,1] \), and
\item[(c)] \( x - y_0 \in M^\perp \)
\end{enumerate}

are equivalent.

Theorem 2.3.5 has some more consequences. It can be employed in deriving results characterizing best approximations from translates of convex cones and subspaces also.

The following result is for translates of convex cones.

**Theorem 2.5.5.** Let \( X \) be a normed space, \( C \) a convex cone in \( X \), \( z \in X \), and \( K = z + C \). Suppose also that \( x \in X \) and \( y_0 \in K \). Then the following statements are equivalent:

\begin{enumerate}
\item[(a)] \( y_0 \in P_K(x) \);
\item[(b)] \( [y'|x - y_0 - \lambda(z + y - y_0)] \leq 0 \text{ and } [y_0 - z|x - y_0 - \lambda(z + y - y_0)] = 0 \text{ for all } y', y \in C \text{ and all } \lambda \in [0,1] \);
\item[(c)] \( [y|x - y_0] \leq 0 \text{ and } [y_0 - z|x - y_0] = 0 \text{ for all } y \in C \).
\end{enumerate}

**Proof.** Being a translate of the convex cone \( C \) in \( X \), \( K = z + C \) is a convex set in \( X \). In fact, if \( k_1 = z + c_1 \) and \( k_2 = z + c_2 \) are in \( K = z + C \), where \( c_1, c_2 \in C \), then for every \( t \in [0,1] \), we have

\[ tk_1 + (1-t)k_2 = tz + (1-t)z + tc_1 + (1-t)c_2 \in z + C = K. \]
Hence by the equivalence of statements (i) and (ii) of Theorem 2.3.5, we have
\[
\begin{align*}
y_0 \in P_K(x) & \iff x - y_0 - \lambda(y' - y_0) \in (K - y_0)\circ \text{ for all } y' \in K \text{ and } \lambda \in [0, 1] \\
& \iff x - y_0 - \lambda(z + y - y_0) \in (z + C - y_0)\circ \text{ for all } y \in C \text{ and } \lambda \in [0, 1] \\
& \iff x - y_0 - \lambda(z + y - y_0) \in (C - (y_0 - z))\circ \text{ for all } y \in C \text{ and } \lambda \in [0, 1] \\
& \iff x - y_0 - \lambda(z + y - y_0) \in C^\circ \cap (y_0 - z)^\perp \text{ for all } y \in C \text{ and } \lambda \in [0, 1], \text{ by Theorem 2.4.5(a), since } y_0 - z \in C \\
& \iff [y' | x - y_0 - \lambda(z + y - y_0)] \leq 0 \text{ and } \\
& [y_0 - z | x - y_0 - \lambda(z + y - y_0)] = 0 \text{ for all } y', y \in C \text{ and } \lambda \in [0, 1].
\end{align*}
\]
Hence (a)\(\Leftrightarrow\)(b).

Similarly, by the equivalence of statements (i) and (iii) of Theorem 2.3.5, it follows that (a)\(\Leftrightarrow\)(c). Hence a)\(\Leftrightarrow\)(b)\(\Leftrightarrow\)(c). \qed

**Remark 2.5.6.** Purely set theoretically, the above theorem can be stated as follows. Under the hypothesis of Theorem 2.5.5, the statements

(a) \(y_0 \in P_K(x)\),  
(b) \(x - y_0 - \lambda(z + y - y_0) \in C^\circ \cap (y_0 - z)^\perp \) for all \(y \in C\) and all \(\lambda \in [0, 1]\), and
(c) \(x - y_0 \in C^\circ \cap (y_0 - z)^\perp\)

are equivalent.

The next result characterizes best approximations from translates of subspaces.
Theorem 2.5.7. Let $X$ be a normed space, $M$ a subspace of $X$, $z \in X$, and $K = z + M$. Suppose also that $x \in X$ and $y_0 \in K$. Then the following statements are equivalent.

(a) $y_0 \in P_K(x)$;

(b) $[y'|x - y_0 - \lambda(z + y - y_0)] = 0$ for all $y'$, $y \in M$ and all $\lambda \in [0, 1]$;

(c) $[y|x - y_0] = 0$ for all $y \in M$.

Proof. $K = z + M$, being a translate of the subspace $M$ of $X$, is a convex set in $X$. Indeed, if $k_1 = z + m_1$ and $k_2 = z + m_2$ are in $K = z + M$, where $m_1, m_2 \in M$, then for every $t \in [0, 1]$, we have

$$tk_1 + (1 - t)k_2 = tz + (1 - t)z + tm_1 + (1 - t)m_2 \in z + M = K.$$  

Hence by the equivalence of statements (i) and (ii) of Theorem 2.3.5, we have

$$y_0 \in P_K(x) \iff x - y_0 - \lambda(y' - y_0) \in (K - y_0)^\circ \text{ for all } y' \in K \text{ and all } \lambda \in [0, 1],$$

$$\iff x - y_0 - \lambda(z + y - y_0) \in (z + M - y_0)^\circ \text{ for all } y \in M \text{ and all } \lambda \in [0, 1],$$

$$\iff x - y_0 - \lambda(z + y - y_0) \in (M - (y_0 - z))^\circ \text{ for all } y \in M \text{ and all } \lambda \in [0, 1],$$

$$\iff x - y_0 - \lambda(z + y - y_0) \in M^\circ \text{ for all } y \in M \text{ and all } \lambda \in [0, 1],$$

$$\iff x - y_0 - \lambda(z + y - y_0) \in M^\perp \text{ for all } y \in M \text{ and all } \lambda \in [0, 1],$$

by theorem 2.4.5 (b)

$$\iff [y'|x - y_0 - \lambda(z + y - y_0)] = 0 \text{ for all } y', y \in M \text{ and all } \lambda \in [0, 1].$$

Hence (a)$\iff$(b).

Similarly, by the equivalence of statements (i) and (iii) of Theorem 2.3.5, it follows that (a)$\iff$(c). Hence (a)$\iff$(b)$\iff$(c).
Remark 2.5.8. A restatement of the above theorem in terms of orthogonal complements is given below. Under the hypothesis of Theorem 2.5.7, the statements

(a) \( y_0 \in P_K(x) \),
(b) \( x - y_0 - \lambda(z + y - y_0) \in M^\perp \) for all \( y \in M \) and all \( \lambda \in [0, 1] \), and
(c) \( x - y_0 \in M^\perp \)

are equivalent.

Remark 2.5.9. Among the four results characterizing best approximations seen so far in this section, namely Theorems 2.5.1, 2.5.3, 2.5.5 and 2.5.7, the one which is in the most general setting is Theorem 2.5.5 characterizing best approximations from translates \( K \) of convex cones \( C \) by \( z \). We observe that, instead of proving each of these results separately, all of these can be deduced directly from an equivalent version of Theorem 2.5.5 in terms of dual cones, which is obtained from Theorem 2.3.5.

We have already seen some results characterizing best approximations from convex sets, in particular from convex cones, subspaces and their translates in this chapter. We conclude our discussions in this chapter with an illustration of one of those characterizations. As a typical case, we consider the illustration of Theorem 2.5.3 characterizing best approximations from subspaces in the following example.

Example 2.5.10. Consider the real normed space \( X = (\mathbb{R}^3, \| \cdot \|_1) \), where

\[
\|x\|_1 = \sum_{k=1}^{3} |x_k| \text{ for } x = (x_1, x_2, x_3) \in X,
\]

equipped with the semi-inner product

\[
[x|y] = \|y\|_1 \sum_{k=1}^{3} \frac{x_k y_k}{|y_k|}, \quad x, y \in X
\]

that generates the norm \( \| \cdot \|_1 \). Consider the subspace

\[
M = \{ x = (x_1, 0, x_2) \in X : x_1, x_2 \in \mathbb{R} \}
\]
of \( X \). Let \( x = (1, 2, 3) \in X \). Then

\[
d(x, M) = \inf \{ \|x - m\|_1 : m \in M \} = \inf \{ |1 - m_1| + |2 - 0| + |3 - m_2| : m_1, m_2 \in \mathbb{R} \} = 2 \text{ at } (1, 0, 3) \in M.
\]

Let \( y_0 = (1, 0, 3) \). Then

\[
\|x - y_0\|_1 = |1 - 1| + |2 - 0| + |3 - 3| = 2.
\]

Thus \( \|x - y_0\|_1 = d(x, M) \), and hence \( y_0 = (1, 0, 3) \in P_M(x) \). Then for every \( m = (m_1, 0, m_2) \in M \), we have

\[
[m|x - y_0| = [(m_1, 0, m_2) | (0, 2, 0)] = \|(0, 2, 0)\|_1 \cdot 0 = 0,
\]

so that \( x - y_0 \in M^\perp \). Now let \( y_0 = (y_1, 0, y_2) \in M \). Suppose that \( x - y_0 \in M^\perp \). Then

\[
0 = [m|x - y_0| \text{ for every } m \in M.] = [(m_1, 0, m_2) | (1 - y_1, 2, 3 - y_2)] \text{ for every } m_1, m_2 \in \mathbb{R} = \|(1 - y_1, 2, 3 - y_2)\|_1 \left[ \frac{m_1 (1 - y_1)}{|1 - y_1|} + \frac{m_2 (3 - y_2)}{|3 - y_2|} \right] \text{ for every } m_1, m_2 \in \mathbb{R}.
\]

This implies that \( y_1 = 1 \) and \( y_2 = 3 \), so that \( y_0 = (1, 0, 3) \). Hence \( y_0 \in P_M(x) \).