1998

A study of optimization in Hilbert Space

John Awunganyi

A STUDY OF OPTIMIZATION IN HILBERT SPACE

A Project

Presented to the

Faculty of

California State University,

San Bernardino

In Partial Fulfillment

of the Requirements for the Degree

Master of Arts

in

Mathematics

---

by

John Awunganyi

June 1998

A STUDY OF OPTIMIZATION IN HILBERT SPACE

A Project

Presented to the

Faculty of

California State University,

San Bernardino

---

by

John Awunganyi

June 1998

Approved by:

[redacted] 6/9/98

Dr. Chetan Prakash, Chair, Mathematics — Date

[redacted]

Dr. Wenxang Wang

[redacted]

Dr. Christopher Freiling

[redacted]

Dr. Terry Hallett, Graduate Coordinator

[redacted]

Dr. Peter Williams, Department Chair

ABSTRACT

The primary objective of this project is to demonstrate that a certain field of optimization can be effectively unified by a few geometric principles of complete normed linear space theory. By employing these principles, important and complex finite - dimensional problems can be interpreted and solved by methods springing from geometric insight. Concepts such as distance, orthogonality, and convexity play a fundamental and indispensable role in this development. Viewed in these terms, seemingly diverse problems and techniques often are found to be closely related.

ACKNOWLEDGEMENTS

My thanks to Dr. Chetan Prakash who not only did his best in making me proficient in mathematical techniques but also showed me how to use these techniques in practical situations. His direction helped make this project plain, direct and readable. I would also like to thank my friends and colleagues Donald Gallaher and Mathias Vheru for their inspiration during the completion of this project. Finally special thanks to my wife, Marvi, for her support during the long hours of research and writing that went into this project.

To My Mother,

Mary Mbonganu Nkemasong

TABLE OF CONTENTS

CHAPTER THREE: APPLICATIONS

CHAPTER FOUR: OPTIMIZATION IN THE BANACH SPACE

Chapter 1

PRELIMINARIES

It is the purpose of this section to explain certain notations, definitions and theorems that shall be used throughout this project. This section does not pretend to be complete, it goes just far enough to establish the connection between the application in question and the basic ideas of functional analysis. This section, therefore, leaves important issues untreated.

Vector Spaces

Definition: A nonempty set V is said to be a **vector space** over a field F, if V is an abelian group under an operation, denoted $+$, and for every $\alpha \in$ F, $x \in V$ there is an element, written αx, in V subject to the following

1. $\alpha(x+y) = \alpha x + \alpha y$
2. $(\alpha + \beta)x = \alpha x + \beta x$
3. $\alpha(\beta x) = (\alpha\beta)x$
4. $1x = x$

for all $\alpha,\ \beta \in F,\ x,\ y \in V$ and 1 represents the unit element of F under multiplication.

Linear dependence

A set S of vectors is said to be **linearly dependent** if to each finite subset $\{x_i\}_{i=1}^n$ there is corresponding a set of scalars $\{\alpha_i\}_{i=1}^n$, not all zero such that

$$\sum_{i=1}^n \alpha_i x_i = 0.$$

A set which is not linearly dependent is said to be **linearly independent**.

Dimensionality of a Vector Space

A vector space is said to be n-dimensional if it contains n-linearly independent vectors and every set with more than n vectors is linearly dependent. A vector space is called infinite-dimensional if there exists an arbitrarily large linearly independent set in the space. In this project we will consider only those vectors spaces with countable dimension.

If an arbitrary vector **x** in V can be represented as a linear combination of a set $\{x_i\}$ in V and scalar $\{\alpha_i\}$ as

$$\mathrm{x} = \sum_{i=1}^{n} \alpha_i x_i .$$

then $\{x_i\}$ is said to **span** the vector space V. A linearly independent set of vectors $\{x_i\}$ that spans a vector space V is called a **basis** for V.

Subspaces and Linear Varieties

Definition. A nonempty subset A of a vector space X is called a **subspace** of X if every vector of the form $\alpha x + \beta y \in$ A whenever x any y are both in A and α, β are any scalars.

Since a subspace is assumed to be nonempty it must contain at least one element. By definition it must contain the zero element. So we can say quite unequivocally that every subspace must contain the null vector. The simplest subspace is the space with the sole element $\{0\}$.

Theorem: Let A and B be subspaces of a vector space X. Then the intersection, A $\cap$ B, of A and B is also a subspace of X.

Proof: Since A and B are subspaces of A and B, it follows that $0 \in A$ and $0 \in B$. Therefore A $\cap$ B is nonempty. Let $x, y \in$ A $\cap$ B, then $x, y \in$ A and $x, y \in$ B. For any scalars α,β the vector $\alpha x + \beta y \in$ A and $\alpha x + \beta y \in$ B since A and B are both subspaces. Therefore $\alpha x + \beta y \in$ A $\cap$ B.

In general this theorem can be extended to any arbitrary number of vectors spaces. We state the extension of this theorem.

Theorem: Let B_α, $\alpha \in I$ be subspaces of a vector space X. Then their arbitrary intersection $\underset{\alpha\in I}{\cap B_\alpha}$ is also subspace of X.

Definition: The sum of two subsets A and B in a vector space, denoted $\mathrm{A}+\mathrm{B}$, consists of all vectors of the form $a+b$ where $a \in A$ and $b \in B$. In other words

$$\mathrm{A}+\mathrm{B} = \{a+b \mid a \in A, b \in B\}.$$

In some literature the word **joint** is used instead of sum and it is sometimes denoted by the lattice symbol $\bigvee$. In this notational parlance, we write

$$\underset{\alpha\in I}{\bigvee B_\alpha} = \text{sum(joints) of the } \mathrm{B}_\alpha\text{'s}$$

The joint $\bigvee B_\alpha$ of a family of subsets is the smallest vector space containing all of them.

Theorem: Let A and B be subspaces of a vector space X. Then their sum $\mathrm{A}+\mathrm{B}$ is a subspace and is equal to their joint.

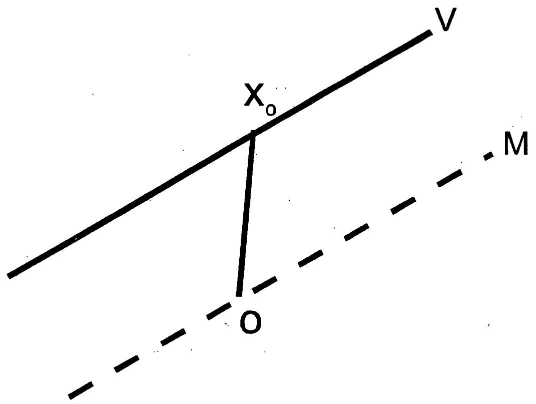
Proof: Since A and B are subspaces of A and B, it follows that $0 \in A$ and $0 \in B$. This implies that $\mathrm{A}+\mathrm{B}$ is nonempty. Suppose x, y are vectors in $\mathrm{A}+\mathrm{B}$. There are vectors a_1, a_2 in A and vectors b_1, b_2 in B such that $x = a_1 + b_1$ and $y = a_2 + b_2$. Given any scalars α, β we can write $\alpha x + \beta y$ as $\alpha(a_1 + b_1) + \beta(a_2 + b_2)$
$= (\alpha a_1 + \beta a_2) + (\alpha b_1 + \beta b_2)$. This shows that $\alpha x + \beta y$ can be expressed as sum of vectors in the subspace A and in the subspace B.

□

Definition: Suppose S is a subset of a vector space X. The set [S] called the **subspace generated by S**, consist of all vectors in X, which are linear combinations of vectors in S.

Definition: A translation of a subspace, M by a fixed vector x_0 is said to be a **linear variety** or **affine subspace.** A subspace is a linear variety if it is the sum of a subspace and a vector.

A linear variety V is usually written as $V = x_0 + M$ where M is a subspace. In this representation M is unique, but any vector in V can serve as x_0. This is illustrated below.



A linear variety

If we are given a subset S, we can easily construct the **smallest** linear variety containing S.

Definition: Let S be a nonempty subset of a vector space X. **The linear variety generated by S,** denoted $v(S)$ is defined as the intersection of all linear varieties in X that contain S.

Convexity and Cones

There is no topic that is responsible for more results in this project than convexity and generalizes many of the useful property of subspaces and linear varieties.

Definition: A set K in a linear subspace is said to be **convex** if given x_i and x_j in K, all points of the form $\alpha x_i + (1-\alpha)x_j$ is also in K if $0 \leq \alpha \leq 1$.

This definition merely says that given two points in a convex set, K, the line segment between them is wholly in K.

Here are some important relations regarding convex sets. As elementary as they may be, they play an important role in proofs involving convex sets.

Theorem: Let K and G be convex sets in a vector space. Then the following are true

a) $\alpha K = \{x : x = \alpha k, k \in K\}$ is convex for any scalar α.

b) $K + G$ is convex - the sum of two convex sets is convex.

Theorem: Let $\mathcal{C}$ be an arbitrary collection of convex sets. Then $\cap_{K \in \mathcal{C}} K$ is convex.

Proof: Let $C = \cap_{K \in \mathcal{C}} K$. If C is empty, then the theorem is true since by definition $\emptyset$ is convex. Assume x_i, $x_j \in C$ and pick α so that $0 \leq \alpha \leq 1$. Then x_i, $x_j \in K$, $\forall\ K \in \mathcal{C}$, and because K is convex $\alpha x_i + (1-\alpha)x_j \in K$ for all $K \in \mathcal{C}$. Thus $\alpha x_i + (1-\alpha)x_j \in$ C and C is convex.

□

We now consider an interesting aspect of norm in terms of convex set, the notion that any sphere is convex.

Theorem: Any sphere is convex.

Proof: Without loss of generality we consider the unit sphere,

$$Y = \{x \in X : ||x|| < 1\}$$

If $x_0, y_0 \in Y$, then $||x_0|| < 1$ and $||y_0|| < 1$. Now if $\alpha \geq 0$ and $\beta \geq 0$, where $\alpha + \beta = 1$, then $||\alpha x_0 + \beta y_0|| \leq ||\alpha x_0|| + ||\beta y_0|| = \alpha ||x_0|| + \beta ||y_0|| \leq \alpha + \beta = 1$ and thus $\alpha x_0 + \beta y_0 \in Y$

□

Cones

Definition: A set C in a linear vector space is said to be a **cone with vertex at the origin** if $x \in C$ implies that $\alpha x \in C$ for all $\alpha \geq 0$. A **convex cone** is a set which is both convex and a cone.

Metric and Normed Linear Spaces

Metric Spaces

Definition: A metric is a set X and a real valued function $d(\ ,\ )$ on X × X which satisfies:

$i)\ d(x,y) > 0$ and $d(x,y) = 0 \Leftrightarrow x = y.$ (positive definiteness)

$ii)\ d(x,y) = d(y,x)$ (symmetry)

$iii)\ d(x,z) \leq d(x,y) + d(y,z)$ (triangle law)

Since a metric space is a set X together with a metric function $d(\ ,\ )$, in general a given set M can be made into a metric space in several different ways by using different metric functions. One such metric on R^n is the so-called 'usual metric' which is defined as $d(x,y) = |x - y|$. In this case, the above three properties simply reflect familiar features of the absolute value or the length function on R^n

Normed Linear Spaces

A vector space that is of particular interest in functional analysis and its application is the **normed linear** space. Such a space come equipped with the topological concepts of openness, closure, convergence and completeness upon introducing the concept of distance on it.

Definition: **A normed linear space** is a vector space X on which there is defined a real valued function which maps each element x in X into a real number $||x||$, called the **norm** of x. The norm respects the following axioms:

a) $||x|| \geq 0$ for all $x \in X$, $||x|| = 0$ if and only if x is a null vector (positive definiteness)

b) $||x + y|| \leq ||x|| + ||y||$ for each $x, y \in X$ (symmetry)

c) $||\alpha x|| = |\alpha|||x||$ for all scalars α and for each $x \in X$. (positive homogeneity)

This is just an abstraction of the usual concept of a metric or length. In a normed linear space, the formula $d(x, y) = ||x - y||$ is a metric. We shall prove this statement by examining the properties (i) to (iii) of a metric space stated above:

$i)$ $d(x, x) = ||x - x|| = ||0 \cdot x|| = |0|||x|| = 0$, If $x \neq y$, then $x - y \neq 0$, so that $d(x, y) = ||x - y|| > 0$.

$ii)$ $d(x, y) = ||x - y|| = ||(-1)(y - x)|| = |-1|||y - x|| = ||y - x|| = d(y, x)$.

$iii)$ $d(x, y) = ||x - y|| = ||x - z + z - y|| \leq ||x - z|| + ||z - y||$

$= d(x, z) + d(z, y)$.

Thus all normed linear spaces are metric spaces

As a direct consequence of the triangular inequality, we state and prove the following result:

Theorem: In a normed linear space X, $\big|||x|| - ||y||\big| \leq ||x - y||$ for any $x, y \in X$.

Proof: $||x|| - ||y|| = ||x - y + y|| - ||y|| \leq ||x - y|| + ||y|| - ||y|| = ||x - y||$ and similarly for $||y|| - ||x||$.

□

Open and Closed Sets

Definition: Let A be a subset of a normed space X. A point $a \in A$ is **an interior point** of A if there is an $\epsilon > 0$ such that all vectors x satisfying $||x - a|| < \epsilon$ are also members of A. The collection of all interior points of A is the **interior** of A which we denote $\mathring{A}$.

Notation: $S(x, \epsilon) = \{y : ||x - y|| < \epsilon$ is the open sphere centered at x with radius ϵ.

Definition: A set S is open if $S = \mathring{S}$.

Definition: A point $x \in X$ is a closure point of a set A if $\forall \epsilon > 0$, there is a point $a \in A$ such that $||x - a|| < \epsilon$. The collection of all closure points of A is the **closure** of A denoted $\overline{A}$. It is clear that $A \in \overline{A}$.

Definition: A set A is closed if $A = \overline{A}$.

Convergence

Definition: In a normed linear space an infinite sequence of vectors $\{x_n\}$ is said to converge to a vector x if the sequence, $\{||x - x_n||\}$ of real numbers converges to zero. We write this as $x_n \to x$.

If $x_n \to x$, then $||x_n|| \to ||x||$ since we have $\Big| ||x_n|| - ||x|| \Big| \leq ||x_n - x|| \to 0$ as $n \to \infty$.

If a sequence converges this limit is unique, since if $x_n \to x$, and $x_n \to y$, then

$$||x - y|| = ||x - x_n + x_n - y|| \leq ||x - x_n|| + ||x_n - y|| \to 0.$$

This can only happen if $x = y$.

Definition: A sequence $\{x_n\}$ in a normed space is said to be a **Cauchy sequence** if $||x_n - x_m|| \to 0$ as $n, m \to \infty$; that is given any $\epsilon > 0$, there is an integer N such that $||x_n - x_m|| < 0$ for all $n, m > N$.

In a normed space, every convergent sequence is a Cauchy sequence since, if $x_n \to x$, then

$$||x_n - x_m|| = ||x_n - x + x - x_m|| \leq ||x_n - x|| + ||x - x_m|| \to 0.$$

We recall from analysis however that a Cauchy sequence may not be convergent. We should also take note of the fact that all Cauchy sequences are bounded. Normed spaces in which every Cauchy sequence has a limit and hence convergent is said to be complete.

Definition: A normed linear vector space X is **complete** if every Cauchy sequence from X has a limit in X. A complete normed linear vector space is called a **Banach space.**

We recall the following fact from analysis:

Theorem: A set F is closed if and only if every convergent sequence with elements in F has its limit in F.

We note that in a finite dimensional linear space, every subspace is automatically closed. This is however, not true for any infinite dimensional space, the proof of which requires the Axiom of Choice.

Transformations and Continuity

Definition: Let X and Y be linear spaces and let D be a subset of X. A rule which associates with every element $x \in D$ and element $y \in Y$ is said to be a transformation from X to Y with domain D. If y corresponds to x under T, we write $y = Tx$.

Definition: A transformation from a vector space X into the space of real or complex scalars is said to be a **functional** on X.

In this project I shall use mostly real-valued functionals, since optimization consists of selecting a vector to minimize or maximize a given functional.

Definition: A transformation T mapping a vector space X into a vector space Y with domain D is said to be **linear** if for every $x_1, x_2 \in D$ and all scalars α_1, α_2 we have $T(\alpha_1 x_1 + \alpha_2 x_2) = \alpha_1 T(x_1) + \alpha_2 T(x_2)$

We recall from analysis that

Definition: A transformation T mapping a normed linear space X in to a normed space Y is **continuous** at $x_o \in$ X if for every $\epsilon > 0$ there is a $\delta > 0$ such that $|x - x_o| < \delta$ implies that $|T(x) - T(x_o)| < \epsilon$

Linear Functionals and Normed Dual

We recall that a functional f on a vector space X is linear if for any two vectors x, $y \in X$, and any two scalars α, β the following always hold:

$f(\alpha x + \beta y) = \alpha f(x) + \beta f(y)$.

Theorem: If a linear functional on a normed linear space X is continuous at a single point, then it is continuous throughout X.

Proof: Assume that f is linear and continuous at $x_o \in X$. Let $\{x_n\}$ be a sequence from X converging to an element $x \in X$. By the linearity of f we have$|f(x_n) - f(x)| = |f(x_n - x + x_o) - f(x_o)|$

Observe that $x_n - x + x_o \to x_o$ and since f is continuous at x_o we have $f(x_n - x + x_o) \to f(x_o)$. Because of this we have

$$|f(x_n) - f(x)| = |f(x_n - x + x_o) - f(x_o)| \to |f(x_o) - f(x_o)| \to 0$$

Thus $|f(x_n) - f(x)| \to 0$. This establishes continuity at all points. □

Definition: A linear functional f on a normed space is **bounded** if there is a constant M such that $|f(x)| \leq M\|x\|$ for all $x \in X$. The smallest such constant M is called the norm of X and is denoted by $\|f\| = \inf\{M : |f(x)| \leq M\|x\|$, for all $x \in X\}$.

A word on notation: The norm of a functional can be expressed in several alternative ways. We list some of them below

$$\|f\| = \inf\{M : |f(x)| \leq M\|x\|, \text{ for all } x \in X\}$$

$$= \sup_{x \neq 0} \frac{|f(x)|}{\|x\|} \text{ or}$$

$$= \sup_{|x| \leq 1} |f(x)| \text{ or}$$

$$= \sup_{|x|=1} |f(x)|$$

Theorem: A linear functional on a normed linear space is bounded if and only if it is continuous.

Proof: Suppose first that a linear functional f is bounded. Let M be such that $|f(x)| \leq M\|x\|$ for all $x \in X$. Then if $x_n \to 0$ then $\|x_n\| \to 0$ and we have

$$|f(x_n)| \leq M\|x_n\| \to 0$$

Thus f is continuous at $x = 0$. From the proceeding theorem, it follows that f is continuous everywhere.

Now, assume that f is continuous at $x = 0$. Then there is a $\delta > 0$, such that $|f(x)| \leq M = 1$ for $\|x_n\| \leq \delta$. Since for any $x \neq 0$ in X, $\delta x/\|x_n\|$ has norm equal to δ, we have the following $|f(x)| = \left|f\left(\frac{\delta x}{\|x_n\|}\right)\right| \times \frac{\|x_n\|}{\delta} < \frac{\|x_n\|}{\delta}$

and $M = \frac{1}{\delta}$ serves as a bound for f.

Norm Dual

Definition: Let X be a normed linear vector space. The space of all bounded linear functionals on X is called the **normed dual of X** denoted X*. The norm of an element $f \in X^*$ is

$$\|f\| = \sup_{|x| \leq 1} |f(x)|$$

The value of the linear functional $x^* \in X^*$ at the point $x \in X$ is denoted by $x^*(x)$ or by the more symmetric notation $< x, x^* >$.

There are several duality principles in optimization theory that relate a problem expressed in terms of vectors to a problem expressed in terms of hyperplanes in the space. Many of the duality principles are based on familiar geometric principles. The shortest distance from a point to a convex set is equal to the maximum of the distances from a point to a hyperplanes separating the point from the convex set. Thus the original minimization over vectors can be converted to maximization over hyperplanes. This is the power afforded by the duality principle - the ability to work in a different space.

Chapter 2

HILBERT SPACES

Inner Product Spaces

Definition: Let X be a a complex vector space. A mapping

$$(.,.) : X \times X \rightarrow \mathbb{C}$$

is called **an inner product** in X if for any $x, y, z \in X$ and $\alpha,\ \beta \in \mathbb{C}$ the following conditions are satisfied.

a) $(x,y) = \overline{(y,x)}$ (conjugate symmetry)

b) $(\alpha x + \beta y, z) = \alpha(x,z) + \beta(y,z)$ (linearity in first part)

c) $(x,x) \geq 0$, $(x,x) = 0$ implies $x = 0$. (positive definiteness)

A vector space with an inner product is called **an inner product space**.

Norm in an inner product Space

An **inner product space** is a vector space with an inner product. It turns out that every inner product space is also a normed space with the norm defined by

$$\|x\| = (x,x)^{\frac{1}{2}}$$

This function is always non-negative. Condition (c) above implies that $\|x\| = 0$ if and only if $x = 0$. Moreover

$$\|\alpha x\| = (\alpha x, \alpha x)^{\frac{1}{2}} = (\alpha\overline{\alpha})^{\frac{1}{2}}(x,x)^{\frac{1}{2}} = |\alpha|\|x\|.$$

For this function to be a norm we need to also prove the triangle inequality. This calls for an intermediate result, the so called Schwarz's inequality. I will state this result without proof.

Lemma: (**Schwarz's Inequality**) For any two elements x and y of an inner product space we have

$$|(x,y)| \leq \|x\|\|y\|$$

The equality $|(x, y)| = ||x|| ||y||$ holds if and only if x and y are linearly dependent.

Triangle Inequality and Parallelogram Law

The Triangle Inequality: For any two elements x and y of an inner product space we have

$$||x + y|| \leq ||x|| + ||y||$$

Proof: $||x + y||^2 = (x + y, x + y) = (x, x) + (x, y) + (y, x) + (y, y)$

$$= (x, x) + 2Re(x, y) + (y, y)$$

$$\leq (x, x) + 2|(x, y)| + (y, y)$$

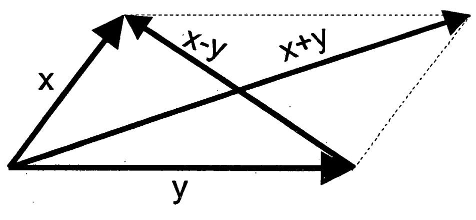
$$\leq ||x||^2 + 2|(x, y)| + ||y||^2$$

$$\leq ||x||^2 + 2||x|| ||y|| + ||y||^2$$

$$\leq (||x|| + ||y||)^2 .$$

Taking the square root of both sides gives the result. □

Though every inner product space is a normed space the converse is not always true. A norm is an inner product space if and only if it satisfies the parallelogram law. This law states that the sum of the squares of the lengths of the diagonals of a parallelogram is equal to twice the sum of the squares of two adjacent sides. This fact is illustrated below.



The parallelogram law

The Parallelogram Law: For any two elements x and y of an inner product space we have

$$\|x+y\|^2 + \|x-y\|^2 = 2(\|x\|^2 + \|y\|^2)$$

Proof: We have

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

and hence

$$\|x+y\|^2 = \|x\|^2 + (x,y) + (y,x) + \|y\|^2 \qquad (1)$$

Now replace y by $-y$ in the above relation we have

$$\|x-y\|^2 = \|x\|^2 - (x,y) - (y,x) + \|y\|^2 \qquad (2)$$

Adding (1) and (2) we have

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

Definition: A complete inner product space is called a **Hilbert space.**

Example: Consider the space $C_{[0,\frac{\pi}{2}]}$ of all continuous functions on the interval $[0,\frac{\pi}{2}]$ with $x(t) = \cos(t)$ and $y(t) = \sin(t)$. This space is not a Hilbert space.

Solution: All we have to do is check if it satisfies the parallelogram law.

$$\|x\| = \|y\| = 1$$

and

$$\|x+y\| = \max_{0\le t\le \frac{\pi}{2}} |\cos(t) + \sin(t)| = \sqrt{2}$$

and

$$\|x-y\| = \max_{0\le t\le \frac{\pi}{2}} |\cos(t) - \sin(t)| = 1$$

Therefore $\|x+y\|^2 + \|x-y\|^2 \neq 2\|x\|^2 + 2\|y\|^2$ since $1+\sqrt{2} \neq 4$

It follows that $C_{[0,\frac{\pi}{2}]}$ cannot be generated by any inner product, that is $C_{[0,\frac{\pi}{2}]}$ fails to be Hilbert.

One of the most important consequences of having the inner product space is the possibility of defining orthogonality of vectors. This makes the theory of Hilbert spaces so different from the other norm spaces.

Definition: Two vectors x and y are said to be **orthogonal** (denoted $x \perp y$) if $(x, y) = 0$.

The Pythagorean Formula: For any pair of orthogonal vectors x and y we have

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

Examples of Hilbert Spaces

Some well known examples of Hilbert spaces are $\mathbb{R}^n$, $\mathbb{C}^n$, $L^2(\mathbb{R})$, $L^2(\mathbb{R}^n)$, and l^2. We prove that the latter is a Hilbert space.

Example: l^2 is a complete inner product space and hence a Hilbert space.

Solution: We recall that l^2 is a vector space with the algebraic operations defined as usual in connection with sequences, that is,

$$(\xi_1, \xi_2, \xi_3, \ldots) + (\eta_1, \eta_2, \eta_3, \ldots) = (\xi_1 + \eta_1, \xi_2 + \eta_2, \ldots)$$

$$\alpha(\xi_1, \xi_2, \xi_3, \ldots) = (\alpha\xi_1, \alpha\xi_2, \alpha\xi_3, \ldots)$$

The inner product here is defined by

$$(x, y) = \sum_{i=1}^{\infty} \xi_i \overline{\eta}_i$$

and the norm defined by

$$\|x\| = (x, x)^{\frac{1}{2}} = \left(\sum_{i=1}^{\infty} |\xi_i|^2\right)^{\frac{1}{2}}$$

In fact, if $x = \{\xi_i\}_1^\infty \in l^2$ any $y = \{\eta_i\}^\infty \in l^2$, then by the Minkowski inequality for sums we have

$$(\sum_{i=1}^{\infty}|\xi_i + \eta_i|^2)^{\frac{1}{2}} \leq (\sum_{i=1}^{\infty}|\xi_i|^2)^{\frac{1}{2}} + (\sum_{i=1}^{\infty}|\eta_i|^2)^{\frac{1}{2}}$$

Since the right hand side is finite , so is the left hand side and implies that $x + y \in l^2$. Also $\alpha x \in l^2$.

By the Cauchy-Schwarz inequality we have

$$\sum_{i=1}^{\infty}|\xi_i \overline{\eta}_i| \leq \left(\sum_{i=1}^{\infty}|\xi_i|^2\right)^{\frac{1}{2}}\left(\sum_{i=1}^{\infty}|\eta_i|^2\right)^{\frac{1}{2}} = \|x\|\|y\|$$

The sequence is bounded by $\|x\|\|y\|$ and hence converges.

To complete the proof, we need to show completeness.

Let $x_m = \{\xi_i^m\}_1^\infty$ be a Cauchy sequence in the space l^2. Then for every $\epsilon > 0$, there is an N such that if $m, n > N$, we have

(1) $$\|x_m - x_n\| = d(x_m, x_n) = \left(\sum_{j=1}^{\infty}\left|\xi_j^m - \xi_j^n\right|^2\right)^{\frac{1}{2}} < \epsilon$$

It follows that for every $j = 1, 2, 3, \ldots$ we have

(2) $$\left|\xi_j^m - \xi_j^n\right| < \epsilon \qquad (m.n > N),$$

For any fixed j, we see from (2) that $(\xi_1^m, \xi_2^m, \xi_3^m \ldots)$ is a Cauchy sequence of numbers. It converges since **R** or $\mathbb{C}$ is complete. Let $\xi_j^m \rightarrow \xi_j$ as m $\rightarrow \infty$. Using these limits we define $x = (\xi_1, \xi_2, \xi_3, \ldots)$ and show that $x \in l^2$ and $x_m \rightarrow x$.

From (1) we have for all $m, n > N$ and for any $k < \infty$,

$$\sum_{j=1}^{k}\left|\xi_j^m - \xi_j^n\right|^2 < \epsilon^2$$

Letting $n \rightarrow \infty$ we obtain for $m > N$

$$\sum_{j=1}^{k}\left|\xi_j^m - \xi_j\right|^2 < \epsilon^2 \qquad \forall k$$

We may now let $k \to \infty$, then for $m > N$

(3) $$\sum_{j=1}^{\infty} \left|\xi_j^m - \xi_j\right|^2 < \epsilon^2 < \infty$$

This shows that $x_m - x = (\xi_j^m - \xi_j) \in l^2$. Since $x_m \in l^2$, it follows by means of the Minkowski inequality that

$$x = x_m + (x - x_m) \in l^2$$

The inequality in (3) also says that, given $\epsilon > 0$, $\exists N$ such that for $m > N$,

$$\|x_m - x\| = \left(\sum_{j=1}^{\infty}\left|\xi_j^m - \xi_j\right|^2\right)^{\frac{1}{2}} < \epsilon,$$

that is $x_m \to x$. Since $x_m = \{\xi_i^m\}_1^{\infty}$ was an arbitrary Cauchy sequence in l^2, this proves completeness of l^2.

Continuity of the Inner Product

The inner product enjoys the following continuity property which is used extensively in this project.

Theorem: (Continuity of the inner product). Suppose that $x_n \to x$ and $y_n \to y$, in an inner product space, then $(x_n, y_n) \to (x, y)$.

Proof: Since the sequence $\{x_n\}$ is convergent, it is bounded by some number say M. So we can write $\|x_n\| < M$. Now we have

$$|(x_n, y_n) - (x, y)| = |(x_n, y_n) - (x_n, y) + (x_n, y) - (x, y)| \leq |(x_n, y_n - y)| + |(x_n - x, y)|$$

Applying the Cauchy-Schwarz inequality, we obtain

$$|(x_n, y_n) - (x, y)| \leq \|x_n\|\|y_n - y\| + \|x_n - x\|\|y\|$$

Since $\|x_n\| < M$, we have

$$|(x_n, y_n) - (x, y)| \leq \mathrm{M}\|y_n - y\| + \|x_n - x\|\|y\| \to 0$$

and hence $(x_n, y_n) \rightarrow (x, y)$ positive definiteness of absolute value

Orthogonal and Orthonormal Systems

Definition: Let X be any inner product space. A family of S of nonzero vectors in X is called an **orthogonal system** if $x \perp y$ for any two arbitrary distinct elements $x, y \in$ S. If in addition, $\|x\| = 1\ \forall x \in S$, S is called an **orthonormal system.**

Every orthogonal set of non-zero vectors can be normalized. If S is an orthogonal system, then the family

$$S_1 = \left\{ \frac{x}{\|x\|} : x \in S \right\}$$

is an orthonormal system. Both systems span the same subspace . We recall that othonormal systems are linearly independent.

Orthonormal Bases

The Hilbert space $\mathbb{C}^n$ is a finite dimensional vector space. Therefore any element of $\mathbb{C}^n$ can be written uniquely as a finite linear combination of a given sets of basis vectors. It follows that the inner product of two elements of $\mathbb{C}^n$ can be computed if we know the expression of each element as such a linear combination. Conversely, the inner product makes possible a very convenient way of expressing a given vector as a linear combination of basis vectors. We recall that if $x_n \in \mathbb{C}^n$ is the n-tupe

$$x_n = (0, 0, ..., 0, 1, 0, ..., 0),$$

where 1 sits in the n-th place. Then $\{x_1, x_2, ..., x_n\}$ is a basis for $\mathbb{C}^n$. Moreover it is clear that

$$(x_n, x_m) = 1 \text{ if } n = m, \qquad (x_n, x_m) = 0 \quad \text{if } n \neq m \qquad (a)$$

If $\mathbf{x} = (\alpha_1, \alpha_2, ..., \alpha_n) \in \mathbb{C}^n$ then the expression for **x** as a linear combination of the basis vector x_n is

$$\mathbf{x} = \sum_{n=1}^{N} \alpha_n x_n$$

Because of the preceding statement (a) we have (x, x_n) $= \alpha_n$. This quantity is called the **Fourier coefficient**. Thus we can write the proceeding expression for **x** as

$$\mathbf{x} = \sum_{n=1}^{N} \alpha_n x_n = \sum_{n=1}^{N} (\mathbf{x}, x_n) x_n \qquad (b)$$

It natural to ask if the series (b) will converges if n is made to approach infinity. We can answer this question by doing the following problem.

Problem: Let V be an inner product space and let $\{x_n\}_{n=1}^{N}$ be an orthonormal set. Prove that $\left\| x - \sum_{n=1}^{N} c_n x_n \right\|$ is minimized by choosing $c_n = (x_n, x)$

Solution: We know that minimizing $\left\| x - \sum_{n=1}^{N} c_n x_n \right\|$ is the same as minimizing

$$\left\| x - \sum_{n=1}^{N} c_n x_n \right\|^2$$

$$\begin{aligned}
\left\| x - \sum_{n=1}^{N} c_n x_n \right\|^2 &= (x - \sum_{n=1}^{N} c_n x_n, x - \sum_{n=1}^{N} c_n x_n) \\
&= (x, x) - 2(\sum_{n=1}^{N} c_n x_n, x) + \left(\sum_{n=1}^{N} c_n x_n, \sum_{n=1}^{N} c_n x_n \right) \\
&= \|x\|^2 - 2\left(\sum_{n=1}^{N} c_n (x_n, x) \right) + \sum_{n=1}^{N} c_n^2 (x_n, x_n) \\
&= \|x\|^2 - 2\left(\sum_{n=1}^{N} c_n (x_n, x) \right) + \left(\sum_{n=1}^{N} c_n^2 \right)
\end{aligned}$$

At this point, observe that $\left((a-b)^2 = a^2 - 2ab + b^2\right) \Rightarrow -2ab + b^2 = (a-b)^2 - a^2$ so we can write,

$$-2\left(\sum_{n=1}^{N} c_n (x_n, x) \right) + \left(\sum_{n=1}^{N} c_n^2 \right) = \sum_{n=1}^{N} (c_n - (x_n, x))^2 - \sum_{n-1}^{N} (x_n, x)^2$$

and so our inner product becomes

$$\left\| x - \sum_{n=1}^{N} c_n x_n \right\|^2 = \|x\|^2 + \sum_{n=1}^{N} (c_n - (x_n, x))^2 - \sum_{n=1}^{N} (x_n, x)^2. \quad *$$

The quantity (*) will be minimized if $(c_n - (x_n, x)) = 0$ and in this case $c_n = (x_n, x)$. The minimum value is

$$\|x\|^2 - \sum_{n=1}^{N} (x_n, x)^2.$$

Moreover, since $\left\| x - \sum_{n=1}^{N} c_n x_n \right\|^2 = (\|x\|^2 - \sum_{n=1}^{N} (x_n, x)^2) > 0$ it follows that

$$\sum_{n=1}^{N} (x_n, x)^2 < \|x\|^2 \text{ for every } n.$$

Hence the series

$$\sum_{i=1}^{\infty} (x_n, x)^2$$

is convergent.

Definition: An **orthonormal basis** for a Hilbert space H is an orthonormal set S $\subset H$ such that the span of S is dense in H. This means that for any $x \in H$ and any $\epsilon > 0$, there is a y, which is a linear combination of elements of S, such that $\|x - y\| < \epsilon$.

Definition: A Hilbert space is said to be **separable** if there is a sequence $\{x_n\}_1^\infty \subset H$ which is dense in H. This means that for any $x \in H$ and any $\epsilon > 0$, there is an n such that $\|x - x_n\| < \epsilon$.

Theorem: H is a separable Hilbert space if and only if H has an orthonormal basis S, which is finite or countable. Two orthonormal basis in a separable Hilbert space must have the same number of elements.

Proof: Suppose $\{\mathrm{u}_n\}_1^\infty$ is a dense subset of H. Then by the Gram-Schmidt orthonomalization process there is a finite or countable orthonormal set S $= \{x_i\}_1^n$ such that each u_n is a linear combination of elements of S. Thus S is an orthonormal basis. This proves the first part.

On the other hand, suppose S is a finite or countable orthonormal basis for H. Consider the subspace T of all vectors **u** of the form

$$\mathbf{u} = \sum_{i=1}^{N} \alpha_i x_i$$

where N is a rational number, the x_i are in S, and the α_i are rational scalars. It is clear that T is countable so the elements of T may be arranged in a sequence of $\{u_n\}_1^\infty$. If $x \in H$, then $x = \sum_{i=1}^{\infty} \beta_i x_i$. For each i, let α_i be rational with $|\beta_i - \alpha_i| < \frac{1}{2^i}$ then

$$\left| x - \sum_{i=1}^{N} \alpha_i x_i \right| \leq \left(\sum_{i=N+1}^{\infty} |\beta_i - \alpha_i|^2 \right)^{\frac{1}{2}} \leq \epsilon \left(\sum_{i=N+1}^{\infty} 2^{-2i} \right)^{\frac{1}{2}} \leq \tfrac{4}{3}(2^{-(N+1)}$$ The last quantity approaches zero as N approaches infinity. Hence for every $e > 0$, $\exists$ element of T within ϵ of x.

□

Orthogonal Complement and the Projection Theorem

A subspace S of a Hilbert space H is an inner product space with the inner product it inherits from H. If we additionally assume that S is a closed subspace of H, then S is a Hilbert space itself, because a closed subspace of an inner product space is complete.

Definition: Let S be a nonempty subset of a Hilbert space H. An element $x \in H$ is said to be **orthogonal** to S, denoted $x \perp$ S, if $(x, y) = 0$ for every $y \in S$. The set of all elements of H orthogonal to S, denoted $S^{\perp}$, is called the **orthogonal complement of S.** In symbols:

$$S^{\perp} = \{x \in H : x \perp S\}$$

The orthogonal complement of $S^{\perp}$ is denoted by $S^{\perp\perp} = (S^{\perp})^{\perp}$

We can easily observe that for any subset S of a Hilbert space H, the set $S^{\perp}$ is a closed subspace of H. It is a subspace because a linear combination of vectors orthogonal to a set is also orthogonal to the set. It is closed since if $\{x_n\}$ is a convergent sequence from $S^{\perp}$, say $x_n \rightarrow x$, continuity of the inner product implies that $(x_n, y) \rightarrow (x, y)$ for all $y \in S$ and so $x \in S^{\perp}$.

Projection Mapping

Definition: For any closed convex subset S of a Hilbert space H, we can define a mapping H into H by assigning to each element x the element closest to x in S, called the **orthogonal projection** of x into S. If $P(x)$ denotes this mapping, P(x) is not necessary linear but is always continuous and convex.

Theorem: Any othogonal projection is continuous.

Proof: Since P is a projection, each z in the inner product space H can be uniquely represented by $z = x + y$ where $x \in S$ and $y \in S^{\perp}$. By definition of orthogonality we have $x \perp y$. It follows by the Pythagorean Theorem that $\|z\|^2 = \|x\|^2 + \|y\|^2$, so $\|P(z)\|^2 = \|x\|^2 \leq \|z\|^2$. This function is bounded and hence continuous.

□

We now turn our attention to a classical optimization problem and the projection theorem which characterizes its solution. There are two slightly different versions of this theorem: one holds in arbitrary inner product space and the other with a much stronger conclusion, is valid in Hilbert space. The optimization problem that gives rise to this theorem can be stated as follows:

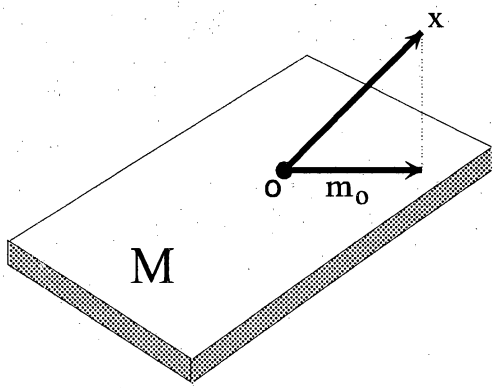
Given a vector x in an inner product space X and a subspace M of X, find a vector $m \in M$ that is closest to x in the sense that it minimizes $\|x - m\|$. Obviously if $x \in M$, then the solution reverts to find the shortest distance between two points. There are three situations we need to consider:

(i) Does such an m exist ?

(ii) If it exists is it unique ?

(iii) What is the solution and how can the solution be characterized?

The three dimensional version of the projection is shown below.



The Projection Theorem

The Projection Theorem (inner product space version)

Let X be an inner product space, M a subspace of X and x and arbitrary vector in X. If there is a vector $m_o \in$ M such that $\|x - m_o\| \leq \|x - m\| \, \forall m \in M$, then m_o is unique. A necessary and sufficient condition that $m_o \in M$ be the unique minimizing vector in M is that the error vector $x - m_o$ be orthogonal to M.

Proof: We first prove that if m_o is a minimizing vector then then the error vector $x - m_o$ is othogonal to M. Suppose there is an $m \in M$ which is not perpendicular to $x - m_o$. To simplify calculations we assume that $\|x\| = 1$ and set $(x - m_o\, , m) = \lambda \neq 0$. Define $m_1 \in M$ as $m_1 = m_o + \lambda m$. Then

$$\|x - m_1\|^2 = \|x - m_o - \lambda m\|^2 = \|x - m_o\|^2 - (x - m_o, \lambda m) - (\lambda m, x - m_o)$$
$$+ |\lambda|^2.$$
$$= \|x - m_o\|^2 - |\lambda|^2 < \|x - m_o\|^2$$

This last statement shows that if $x - m_o$ is not orthogonal to M, then m_o cannot be a minimizing vector. Finally we need to show that if $x - m_o$ is orthogonal to M, then m_o is the unique minimizing vector. For any $m \in M$ and $x \in X$, the Pythagorean theorem gives

$$\|x - m\|^2 = \|x - m_o + m_o - m\|^2 = \|x - m_o\|^2 + \|m_o - m\|^2 \geq \|x - m_o\|^2$$

The above implies that $\|x - m\| > \|x - m_o\|$ if and only if $m \neq m_o$.

In the discussion above, we have shown that if the minimizing vector exists it must be unique and that $x - m_o$ is orthogonal to the subspace M. By making the hypotheses a little stronger, we can guarantee the existence of the of the minimizing vector. This can be achieved by making the subspace M a closed space. This is shown in the following more powerful version.

Projection Theorem (Hilbert space version):

Let H be Hilbert space, M a closed subspace of H. Corresponding to any vector x in H, there is a unique vector $m_o \in$ M such that $\|x - m_o\| \leq \|x - m\| \ \forall m \in M$. A necessary and sufficient condition that $m_o \in M$ be the unique minimizing vector in M is that $x - m_o$ be orthogonal to M.

Proof: The uniqueness and orthogonality is established above in the inner product version of this theorem. All that is needed is the existence of the minimizing vector. If $x \in M$, then $m_o = x$ and we are done. Let assume that $x \notin M$ and define $d = \inf_{m \in M} \|x - m\|$. Our goal is to produce $m_o \in M$ with $\|x - m_o\| = d$. For this purpose, let $\{m_i\}$ be a sequence of vectors in M such that $\|x - m_i\| \to d$. By the parallelogram law, we have the following

$$\|(m_j - x) + (x - m_i)\|^2 + \|(m_j - x) - (x - m_i)\|^2 = 2\|m_j - x\|^2 + 2\|x - m_i\|^2.$$

Rearranging we get,

$$\|m_j - m_i\|^2 = 2\|m_j - x\|^2 + 2\|x - m_i\|^2 - 4\left\|x - \frac{m_i + m_j}{2}\right\|^2$$

Since M is a linear subspace, the vector $\frac{m_i + m_j}{2}$ is in M since $m_i, m_j \in M \,\forall\, i, j$.
By the way we define d, $\left\|x - \frac{m_i + m_j}{2}\right\| \geq d$ and we have

$$\|m_j - m_i\|^2 \leq 2\|m_j - x\|^2 + 2\|x - m_i\|^2 - 4d^2$$

As $i \to \infty$ both $\|m_j - x\|^2$ and $\|x - m_i\|^2$ approaches d^2 and hence

$$\|m_j - m_i\|^2 \leq 2d^2 + 2d^2 - 4d^2 = 0$$

We can now conclude that $\|m_j - m_i\|^2 \to 0$ as $i, j \to \infty$. Therefore the $\{m_i\}$ is a Cauchy sequence, and since M is a closed subspace of a Hilbert space, the sequence $\{m_i\}$ has a limit $m_o \in M$. Since the norm respects the continuity property, it follows that $\|x - m_o\| = d$.

□

Chapter 3

APPLICATIONS

The main purpose of this section is to examine a variety of problems that can be formulated as optimization problems in the Hilbert space by examining certain specific examples. We shall look at instances of Approximation Theory, Game Theory (where we prove the min-max theorem), Control-type Problems and Minimum Distance to a convex set.

Approximation Theory

The motivation behind all approximation problems is the desire to approximate a general mathematical situation by a simpler, more specific form. In this section we shall look at two different situations, viz, the normal equation and the Gram Matrices, and the Fourier Series method of approximation.

Normal Equation and the Gram Matrices

Suppose a given decision maker wants to investigate the following situation: We are given that $y_1, y_2, y_3, ..., y_n$ all belong to some Hilbert space H, generating a subspace $M \subset H$. Given an arbitrary vector $x \in H$ we seek a vector $y \in M$ which is closest to x. Now y can be expressed as a linear combination of the y_i say $y = \alpha_1 y_1 + \alpha_2 y_2 + ... + \alpha_n y_n$, where $\alpha_i \in R, i = 1, 2, ..., n$. The problem now is equivalent to find the α_i such that the quantity

$$\|x - y\| = \|x - (\alpha_1 y_1 + \alpha_2 y_2 + ... + \alpha_n y_n)\|$$

is minimized. The projection theorem can easily be used to solve this problem. According to the projection theorem, the minimizing vector y is the orthogonal projection of x on M.

Another way of putting it is that the difference vector $x - y$ must be orthogonal to each of the y_i. Therefore we can write

$$(x - \alpha_1 y_1 - \alpha_2 y - ... - \alpha_n y, y_i) = 0$$

for $i = 1, 2, 3, ..., n$. Writing this in expanded form and recalling that $(y_i, y_i) = \delta_{ij}$, we have,

$$(y_1, y_1)\alpha_1 + (y_2, y_1)\alpha_2 + ... + (y_n, y_1)\alpha_n = (x, y_1)$$

$$(y_2, y_2)\alpha_1 + (y_2, y_2)\alpha_2 + ... + (y_n, y_2)\alpha_n = (x, y_2)$$

$$\vdots$$

$$(y_1, y_n)\alpha_1 + (y_2, y_n)\alpha_n + ... + (y_n, y_n)\alpha_n = (x, y_n)$$

These n equations are called the the **normal equations** for the minimization problem. Corresponding to the vectors $y_1, y_2, y_3, ..., y_n$, the square $n \times n$ matrix G is

$$G = G(y_1, y_2, y_3, ..., y_n) = \begin{bmatrix} (y_1, y_1) & (y_1, y_2) & \cdots & (y_1, y_n) \\ (y_1, y_1) & & & \\ \vdots & & & \vdots \\ (y_n, y_1) & & \cdots & (y_n, y_n) \end{bmatrix}$$

We recall from Linear Algebra that G$(y_1, y_2, y_3, ..., y_n)$ is the **Gram matrix** of $\{y_1, y_2, y_3, ..., y_n\}$. It is the transpose of the coefficient matrix of the normal equations. The approximation problem will be solved once the normal equations are solved. In order for the normal equations to be solvable the Gram determinant must be nonzero. That is, it must be invertible. This can only happen if the vectors $\{y_1, y_2, y_3, ..., y_n\}$ are linearly independent. Once this fact is established the finding of the minimum distance from x to the subspace M can be found by Cramer's rule. Cramer's rule, which until several years ago was of little practical importance because of the difficulties of evaluating large determinants has now found a new audience because of computers and high speed calculators. With the availability of high speed digital computers it is also easy to find the inverse of the invertible matrix and hence the solution. This method was also avoided in the past because of the difficulty and cumbersomness of finding inverses of large matrices.

We now return our attention to the evaluation of the minimum distance between x and the subspace M by applying the following theorem from Linear Algebra.

Theorem: Let the $y_1, y_2, y_3, \ldots, y_n$ be linearly independent. Let d be the minimum distance from a vector x to the subspace M generated by the y_i's, that is

$$d = \min \|x - \alpha_1 y_1 - \alpha_2 y_2 - \ldots - \alpha_n y_n\|$$

Then

$$d^2 = \frac{\det(G(y_1, y_2, \ldots, y_n, x))}{\det(G(y_1, y_2, \ldots, y_n))}.$$

Proof: If $y \in M$ is a minimizing vector for the distance then $d^2 = \|x - y\|^2 = (x - y, x - y) = (x - y, x) - (x - y, y)$. By the projection theorem, $x - y$ is orthogonal to M and as a result $(x - y, y) = 0$. Therefore,

$$d^2 = (x - y, x) = (x, x) - \alpha_1(y_1, x) - \alpha_2(y_2, x) - \ldots - \alpha_n(y_n, x)$$

Rearranging, $\alpha_1(y_1, x) + \alpha_2(y_2, x) + \ldots + \alpha_n(y_n, x) + d^2 = (x, x)$

The above equation along with the normal equations, yields $n + 1$ linear equations for the $n + 1$ unknowns, $\alpha_1, \ \alpha_2, \ldots, \ \alpha_n$, d^2 which is solvable by Cramer rule. The value of d^2 is

$$d^2 = \frac{\det\begin{bmatrix} (y_1,y_1) & (y_2,y_1) & \dots & (y_n,y_1) & (x,y_1) \\ (y_1,y_2) & & & & \\ \cdot & & & & \cdot \\ \cdot & & & & \cdot \\ \cdot & & & & \cdot \\ (y_1,y_n) & & \dots & (y_n,y_n) & (x,y_n) \\ (y_1,x) & & \dots & (y_n,x) & (x,x) \end{bmatrix}}{\det\begin{bmatrix} (y_1,y_1) & (y_2,y_1) & \dots & (y_n,y_1) & 0 \\ (y_1,y_2) & & & & \\ \cdot & & & & \cdot \\ \cdot & & & & \cdot \\ \cdot & & & & \cdot \\ (y_1,y_n) & & \dots & (y_n,y_n) & 0 \\ (y_1,x) & & \dots & (y_n,x) & 1 \end{bmatrix}} = \frac{\det(G(y_1,y_2,\dots,y_n,x))}{\det(G(y_1,y_2,\dots,y_n))}.$$

□

Fourier Series Approximation

Finding the best approximation to x in the subspace M, where M is generated by orthonormal vectors $x_1, x_2, \dots, x_n$ is a special case of the general approximation problem I discussed above. In this special case, we see immediately that the general approximation problem is trivial because the Gram matrix of the x_i's is simply the identity matrix giving the best approximation to be

$$y = \sum_{i=1}^{n}(x, x_i)x_i = \sum_{i=1}^{n}\alpha_i x_i$$

Our goal in this section is to extended this special approximation problem slightly by considering approximation in a closed subspace generated by an infinite orthonormal system. Before we do that we must recall from analysis the following definition of convergence of an infinite series.

Definition: An **infinite series** of the form $\sum_{i=1}^{\infty} x_i$ is said to converge to an element x in a normed space if the sequence of partial sums $s_n = \sum_{i=1}^{n} x_i$ converges to x; then we write $x = \sum_{i=1}^{\infty} x_i$.

The next theorem establishes the necessary and sufficient condition for an infinite series of orthonormal vectors to converge in a Hilbert space.

Theorem : Let $\{x_i\}$ be an orthonormal sequence in a Hilbert H. A series of the form $\sum_{i=1}^{\infty} \alpha_i x_i$ converges to an element $x \in H \Leftrightarrow \sum_{i=1}^{\infty} |\alpha_i|^2 < \infty$, and in this case $\alpha_i = (x, x_i)$.

Proof: Suppose that $\sum_{i=1}^{\infty} |\alpha_i|^2 < \infty$ and let $s_n = \sum_{i=1}^{n} \alpha_i x_i$, then

$$\|s_n - s_m\|^2 = \left\| \sum_{i=m+1}^{n} \alpha_i x_i \right\|^2 = \sum_{i=m+1}^{n} |\alpha_i|^2 \to 0 \quad \text{as } m, n \to \infty.$$

This implies that $\{s_n\}$ is a Cauchy sequence and because H is complete there is an element $x \in H$ such that $s_n \to x$.
On the other hand, if s_n converges, then it is a Cauchy sequence so $\sum_{i=n+1}^{m} |\alpha_i|^2 \to 0$. Thus $\sum_{i=n+1}^{\infty} |\alpha_i|^2 \to 0$ and $\sum_{i=1}^{\infty} |\alpha_i|^2 < \infty$.

To show the last part we observe that $(s_n, x_i) = \alpha_i$ as soon as $n \geq i$, which by the continuity of inner product implies $(x, x_i) = \alpha_i$.

□

We recall from analysis that $(x, x_i) = \alpha_i$ is called the **Fourier coefficient** of x with respect to the orthonormal basis $\{x_i\}$. The Fourier coefficients and vectors are related by the following

$$\sum_{i=1}^{\infty}|(x, x_i)|^2 \leq \|x\|^2$$

This relation is called **Bessel's inequality** and guarantees that $\sum_{i=1}^{\infty}|(x, x_i)|^2 \leq \infty$. The above theorem also guarantees the fact that $\sum_{i=1}^{\infty}(x, x_i)x_i$ converges to some element. We characterize this element in the next theorem.

Theorem: Let x be an element in a Hilbert space H and suppose $\{x_i\}$ is an orthonormal sequence in H. Then the series

$$\sum_{i=1}^{\infty}(x, x_i)x_i$$

converges to an element y in the closed subspace M generated by the x_i's. The "error" vector $x - y$ is orthogonal to M.

Proof: Convergence is guaranteed by the last theorem and by Bessel's inequality. Since M is closed $y \in M$. The sequence of partial sums $s_n = \sum_{i=1}^{n}(x, x_i)x_i \rightarrow y \in M$. For each j and $n > j$ we have

$$(x - s_n, x_j) = \left(x - \sum_{i=1}^{n}(x, x_i)x_i, x_j\right) = (x, x_j) - (x, x_j) = 0.$$

Therefore by the continuity of the inner product $\lim_{n \to \infty}(x - s_n, x_j) = (x - y, x_j) = 0$ for each j. Thus $x - y$ is orthogonal to the subspace generated by the x_i's. Again using the continuity of the inner product we can conclude that $x - y$ is orthogonal to the closed subspace generated by the x_i's.

□

It is now clear that if a closed subspace generated by the orthonormal set of $\{x_i\}$ is the whole space, then any vector in the Hilbert space H, can be expanded as a series of the x_i's with coefficients equal to the Fourier coefficients (x, x_i). In fact to express every $x \in H$ as the limit of an infinite sequence of the form $\sum_{i=1}^{\infty}\alpha_i x_i$ it is necessary that the closed subspace generated by x_i's be the whole space.

Suppose again that we are given independent, but not necessarily orthonormal vectors $y_1, y_2, y_3, ..., y_n$ generating a subspace M of Hilbert space H and we wish to find the vector $y \in M \subset H$ which minimizes $||x - y||$. This time rather than seeking to obtain y directly as a linear combination of the y_i's by solving the normal equations, we can simply employ the Gram-Schmidt othogonalization procedure and then the Fourier series approximation as above. First we apply the Gram-Schmidt othogonalization procedure to $\{y_1, y_2, y_3, ..., y_n\}$, and obtain the orthonormal set $\{x_1, x_2, \ldots, x_n\}$ generating M. The vector y can then be written in terms of the Fourier coefficients as

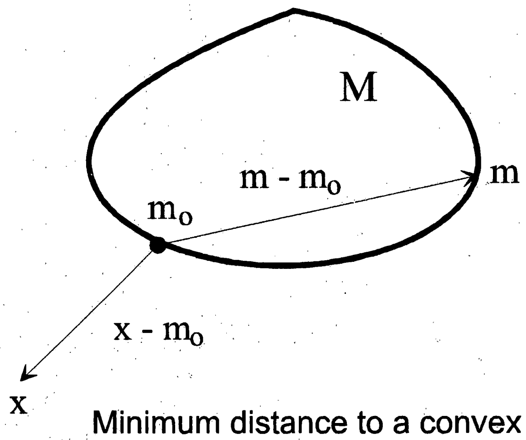
$$y = \sum_{i=1}^{n}(x, x_i)x_i$$

and $(x - y) \perp M$. Thus our original optimization problem can now be solved with relative ease since we have orthonomalized the independent vectors. Since the solution to the approximation problem is equivalent to the solution of the normal equations, we can conclude that the Gram-Schmidt procedure can be interpreted as an algorithm for inverting the Gram matrix. In fact the Gram-Schmidt procedure consists of solving a series of minimum norm approximation problems by the use of the projection theorem. So we can solve the minimum norm approximation on the subspace generated by $\{y_1, y_2, y_3, ..., y_n\}$ by applying the Gram-Schmidt procedure to the sequence $\{y_1, y_2, y_3, ..., y_n, x\}$. The optimal error $x - y$ is found at the last step of the process.

Minimum Distance to a Convex Set

(Closest Point Property)

The closest point to a convex set is of fundamental importance to many approximation problems. The following theorem, concerning the minimization of the norm, is illustrated below and is a direct extension of the proof of the projection theorem.



Minimum distance to a convex set

Theorem: Let x be a vector in a Hilbert space H and let M be a closed convex subset of H. Then there is a unique vector $m_0 \in M$ such that

$$\|x - m_0\| \leq \|x - m\|$$

for all $m \in M$. Furthermore, a necessary and sufficient condition that m_0 be a unique minimizing vector is that $(x - m_0, m - m_o) \leq 0$ for all $m \in M$.

Proof: To show existence, let $\{m_i\}$ be a sequence from M such that

$$\|x - m_i\| \rightarrow d = \inf_{m \in M} \|x - m\|$$

We now apply the parallelogram law to get

$$\|m_i - m_j\|^2 = 2\|m_i - x\|^2 + 2\|m_j - x\|^2 - 4\left\|x - \tfrac{m_i+m_j}{2}\right\|^2$$

Because M is convex, $\frac{m_i+m_j}{2}$ is in M; and hence $\left\|x - \frac{m_i+m_j}{2}\right\|$ must be at least d, and so we have

$$\left\|x - \tfrac{m_i+m_j}{2}\right\| \geq d$$

and therefore

$$\|m_i - m_j\|^2 \leq 2\|m_i - x\|^2 + 2\|m_j - x\|^2 - 4d^2 \rightarrow 4d^2 - 4d^2 = 0.$$

Therefore the sequence $\{m_i\}$ is Cauchy and hence convergent to an element $m_0 \in M$. Using the continuity property of inner product, $\|x - m_0\| = d$.

To prove uniqueness, suppose $m_1 \in M$ with $\|x - m_1\| = d$. The sequence $\{m_i\}$ has $\|x - m_i\| \to d$ so by the above argument, $\{m_i\}$ is Cauchy and convergent. Then we claim that $\frac{m_1 + m_0}{2} \in M$ (by convexity) is closer to x than d. For

$$* \qquad \left\| x - \frac{m_1 + m_0}{2} \right\| = \frac{1}{2}(\|(x - m) + (x - m_0)\|)$$

and this equals $\frac{1}{2}\|x - m_1\| + \frac{1}{2}\|x - m_0\| = d$ only if $x - m_1$ is a multiple of $x - m_0$.

Now $d = \|x - m_1\| = \|x - m_0\|$, so if $x - m_1$ is a multiple: $x - m_1 = \alpha(x - m_0)$, then

$$|\alpha| = 1. \text{ If } \alpha = 1,\ x - m_1 = x - m_0 \Rightarrow m_1 = m_0$$

$$\text{if } \alpha = -1,\ x - m_1 = m_0 - x \Rightarrow x = \frac{m_1 + m_0}{2} \in M, \text{ a contradiction.}$$

Thus if $x - m_0$ is not a multiple, the original quantity (*) is less than d, which say that $\frac{m_1 + m_0}{2} \in$ M is closer than m_0, a contradiction. Hence $m_1 = m_0$.

We now show that if m_0 is the unique minimizing vector in M, then

$$(x - m_0, m_1 - m_0) \leq 0$$

for all $m \in M$. Suppose to the contrary that there is a vector $m_1 \in M$ such that $(x - m_0, m_1 - m_0) = \epsilon > 0$. Pick any vector $m_\alpha \in M$, such that

$m_\alpha = (1 - \alpha)m_0 + \alpha m_1$; $0 \leq \alpha \leq 1$. Since M is convex, each $m_\alpha \in M$. Also

$$\|x - m_\alpha\|^2 = \|(1 - \alpha)(x - m_0) + \alpha(x - m_1)\|^2$$
$$= (1 - \alpha)^2\|x - m_0\|^2 + 2\alpha(1 - \alpha)(x - m_0, x - m) + \alpha^2\|x - m_1\|^2$$

The quantity $\|x - m_\alpha\|^2$ is a differentiable function of α with derivative at $\alpha = 0$ given by

$$\left.\frac{d}{d\alpha}\|x - m_\alpha\|^2\right|_{\alpha=0} = -2\|x - m_0\|^2 + 2(x - m_0, x - m_1)$$
$$= -2(x - m_0, m_1 - m_0) = -2\epsilon < 0.$$

Thus for some positive α, $\|x - m_\alpha\| < \|x - m_0\|$. This contradicts the minimality of m_0. Thus no such m_1 exist.

Conversely, suppose that $m_0 \in M$ is such that $(x - m_0, m - m_0) \leq 0$ for $m \in M$. Then for any $m \in M$, with $m \neq m_0$, we have

$$\|x-m\|^2 = \|x-m_0+m_0-m\|^2$$
$$= \|x-m_0\|^2 + 2(x-m_0, m_0-m) + \||m_0|-m\|^2 > \|x-m_0\|$$

showing that m_0 is a minimizing vector.

□

Problem: As an application of the minimum norm problem, we consider an approximation problem with restriction on the coefficients. Let $\{y_1, y_2, y_3, ..., y_n\}$ be linearly independent vectors in a Hilbert space H. Given $x \in H$, we seek to minimize $\|x - \alpha_1 y_1 - \alpha_2 y_2 - ... - \alpha_n y_n\|$ where we require $\alpha_i \geq 0$ for each i.

Solution: We can reformulate this problem abstractly as that of finding the minimum distance from a point x to the convex cone

$$M = \{y : y = \alpha_1 y_1 + \alpha_2 y_2 + ... + \alpha_n y_n, \alpha_i \geq 0 \text{ for each } i\}$$

M is a closed convex cone and hence, there is a unique minimizing vector. The minimizing vector $\widehat{x} = \alpha_1 y_1 + \alpha_2 y_2 + ... + \alpha_n y_n$ must satisfy

$$(x - \widehat{x}, m - \widehat{x}) \leq 0, \qquad \text{for all } m \in M$$

Setting $m = \widehat{x} + y_i$ leads to

$$(x - \widehat{x}, y_i) \geq 0 \qquad \text{if } \alpha_i > 0$$

and setting $m = \widehat{x} + \alpha_i y_i$ leads to

$$(x - \widehat{x}, y_i) \leq 0 \qquad \text{for } i = 1, 2, 3, \ldots, n$$

with equality $\alpha_i > 0$.

Letting G be the Gram matrix of $\{y_1, y_2, y_3, ..., y_n\}$ and letting $b_i = (x, y_i)$, we obtain the matrix equation:

$(i\ )$ $$G\alpha - b = z$$

for some vector z with component $z_i \geq 0$. We recall from linear algebra that α and b are vectors with component represented by α_i and b_i respectively. Furthermore, $\alpha_i z_i = 0$.

Condition (i) above is analog of the normal equation.

Control Problems

Problems of control are associated with dynamical systems evolving in time. These types of problems usually refer to directed influence on a dynamic system to achieve a desired result. The system itself may be physical such a Sojourner rocket heading for Mars with a rover or a chemical plant processing milk or it may be operational such as a warehouse receiving and filling orders.

Often we seek a feedback in which a decision of current control actions are made continuously in time based on periodic observations of system behavior. One may imagine himself as a controller sitting in a control panel watching meters and turning knobs or in a warehouse ordering new stock based on inventory and predicted demand.

Any control problem might be formulated in a vector space consisting of an optimal control function $u(t)$ defined on an interval $[0, T]$. For the motion of the rocket being propelled vertically the governing equation may be

$$\frac{d^2y}{dt^2} = u(t) - g$$

where y is the vertical displacement, u is the accelerating force, and g is the gravitational force. The optimal control function u is the one which forces $y(T) = h$ while minimizing the fuel expenditure, which we represent in this case by $\int_0^{\mathrm{T}} |u(t)| dt$.

Minimizing a Quadratic Objective

Let us consider an optimal control problem which seek to minimize the quadratic objective given by

$$G = \int_0^T \{x^2(t) + u^2(t)\} dt$$

where x and u are related by the differential equation

(1) $\frac{dx(t)}{dt} = u(t)$; and the initial condition $x(0)$ is given.

We want to reduce x to zero quickly by a suitable application of the control function $u(t)$. The quadratic objective represents a common compromise between a desire to make x small while at the same time maintaining control over $u(t)$. We can begin by replacing equation (1) by the equivalent constraint

(2) $$x(t) = x(0) + \int_0^t u(\tau)d\tau.$$

We are now in a position to formulate the above problem in the Hilbert space

$$H = L_2[0, T] \times L_2[0, T]$$

consisting of the ordered pairs (x, u) of square – integrable functions on $[0, T]$ with the inner natural inner product defined by

$$((x_1, u_1), (x_2, u_2)) = \int_0^T [x_1(t)x_2(t) + u_1(t)u_2(t)]dt;$$

and the corresponding norm is

$$\|(x, u)\|^2 = \int_0^T \left[x^2(t) + u^2(t)\right] dt.$$

We have now defined the norm in the Hilbert space and we recall that the set of elements $(x, u) \in H$ satisfying the constraint (2) is a linear variety $V \in H$. The control problem is now one of finding the element $(x, u) \in V$ having a minimum norm.

If V is closed we have a unique solution in V. To prove that V is closed, let $\{(x_n, u_n)\}$ be a sequence of elements from V converging to an element (x, u). For V to be closed (x, u) must lie in V. Letting $y(t) = x(0) + \int_0^t u(\tau)d\tau$, we must show that $x = y$. Thus by the Cauchy-Schwarz inequality, applied to the functions (1) and $u(t) - u_n(t)$, and integrating from 0 to T we obtain

$$|y(t) - x_n(t)|^2 \leq t\int_0^t |u(\tau) - u_n(\tau)|^2 d\tau \leq T\|u - u_n\|^2$$

and hence $\|y - x_n(t)\| \leq T\|u - u_n\|$. It follows that

$$||y - x|| \leq ||y - x_x|| + ||x_n - x|| \leq T||u - u_n|| + ||x_n - x||$$

The two terms on The right tend to zero as $n \to \infty \Rightarrow x = y$.

Game Theory

Many problems that involve a competitive element can be regarded as games. In the usual formulation involving two players, or two antagonists, there is an objective function whose value depends jointly on the action employ by both players. One player attempts to maximize this objective, and the other attempts to minimize it. Often several problems from almost any area of mathematics can be intermixed to produce a game. Some game theoretic problems are of the pursuer-evader type such as a fighter plane chasing a bomber. Each player has a system he controls but one is trying to maximize the objective (time to intercept for instance) while the other is trying to minimize that objective.

As another example, we consider a problem of advertising or campaigning by two students of California State University, San Bernardino running for Student Body President. Two opposing students, A and B are running and must plan how to allocate advertising resources (A and B dollars respectively) among n distinct departments and groups. We can let x_i and y_i represent, respectively, the resources allocated to department i by candidate A and B. We assume that there are currently a total of u undecided votes in the whole university and u_i representing the number of undecided in department i. According to some determined mathematical model, the number of votes a given candidate received in each department is

$$\frac{\text{(Amount of money spent by a candidate in a department)(Number of undecided in that department)}}{\text{(Total amount spent by the candidates in the department)}}$$

Using this model the number of votes going to candidates A and B from department i will be

$$\frac{x_i u_i}{x_i^2+y_i^2}, \frac{y_i u_i}{x_i^2+y_i^2}$$

respectively. The total difference in votes between the number of votes received by A and B will then be

$$\sum_{i=1}^{n} \frac{x_i - y_i}{x_i^2+y_i^2} u_i .$$

Candidate A will seek to maximize this quantity while Candidate B will seek to minimize it.

Min-Max Theorem of Game Theory

Before we present and prove the min-max theorem we need some background facts . Let X be a normed space and X* its normed dual space. Let A be a fixed subset of X and B a fixed subset of X*. In this game player A selects a vector from the strategy set A while his opponent player B selects a vector from his strategy set B. When both players have selected their respective vectors the quantity $< x, x* >$ is computed and player A pays the amount to player B. Thus A seeks to make his selection so as to minimize $< x, x* >$ while B seeks to maximize $< x, x* >$. Assume for the time being the existence of the quantities

$$Q^{o} = \min_{x \in A} \max_{x* \in B} < x, x* >$$

$$Q_{o} = \max_{x* \in B} \min_{x \in A} < x, x* >$$

Consider first the viewpoint of A in this game. By selecting $x \in A$, he looses no more than $\max_{x* \in B} < x, x* >$, hence by proper choice of x, say x_o , he can be assured of loosing no more than Q^o. On the other hand player B by selecting $x* \in B$, wins at least a $\min_{x \in A} < x, x* >$ so by a judicious choice of $x*$, say x_o* he can be garantee a win of at least Q_o. It follows that $Q_o \leq < x, x* > \leq Q^o$. If $Q^o = Q_o$ we have a draw and there is a well-determined pay-off value for optimal play by both players. The min-max theorem simply states that for approrpriate sets A and B, $Q^o = Q_o$.

We now present the proof of the min-max theorem based on duality. For simplicity this version of the proof is for reflexive spaces that is in spaces where $(X^*)^* = X$. This proof of the min-max theorem make extensive use of the following Fenchel Duality Theorem which we state without proof. Before we state the Theorem we give the following definition.

Definition: In correspondence to a convex functional f defined on a convex set C in a vector space X, we define the convex set $[f, C]$ in $R \times X$ as

$$[f, C] = \{(r, x) \in R \times X : x \in C, f(x) \leq r\}$$

Theorem: (Fenchel Duality Theorem). Assume that f and g are, respectively, convex and concave functionals on the convex sets C and D in a normed space X. Assume that $C \cap D$ contains points in the relative interior of C and D and that either $[f, C]$ or $[g, D]$ has nonempty interior. Suppose further that $\mu = \inf_{x \in C \cap D} \{f(x) - g(x)\}$ is finite. Then

$$\mu = \inf_{x \in C \cap D} \{f(x) - g(x)\} = \max_{x^* \in C^* \cap D^*} \{g^*(x^*) - f^*(x^*)\}$$

where the maximum on the right is achieved by some $x_o^* \in C^* \cap D^*$. If the infimum on the left is achieved by some $x_o \in C \cap D$, then

$$\max_{x \in C} [< x, x_o^* > - f(x)] = < x_o, x_o^* > - f(x_o)$$

and

$$\min_{x \in D} [< x, x_o^* > - g(x)] = < x_o, x_o^* > - g(x_o)$$

Theorem (min-max): Let X be a reflective normed space and let A and B be compact convex sets of X and X*, respectively. Then

$$\min_{x \in A} \max_{x* \in B} < x, x* > = \max_{x* \in B} \min_{x \in A} < x, x* >$$

Proof: Define a functional f on X by

$$f(x) = \max_{x* \in B} < x, x* >$$

The maximum exists for each $x \in X$ since B is compact. f is also continuous and convex on X. We seek an expression for

$$\min_{x \in A} f(x)$$

which exists because of the compactness of A and the continuity of f. We now apply the Fenchel duality theorem with the associations: $f \to f$, $C \to X$, $g \to 0$, and $D \to A$. We have immediately the following associations

(1) $$D^* \to X^*$$

(2) $$g^*(x^*) = \min_{x \in A} < x, x* >$$

We further claim that

(3) $$C^* = B$$

(4) $$f^*(x^*) = 0$$

To prove (3) and (4), let $x_1^* \notin B$ and by using the separating hyperplane theorem, let $x_1 \in X$ and α be such that $< x_1, x_1^* > - < x_1, x^* > > \alpha > 0$ for all $x^* \in B$. Then

$< x_1, x^* > - \max_{x*\in B} \min_{x\in A} < x, x^* >$ can be made arbitrarily large by taking $x = kx_1$ with $k > 0$. Thus

$$\sup_x [< x, x_1^* > - f(x)] = \infty$$

and $x_1^* \notin C^*$.

Conversely, if $x_1^* \in$ B, then $< x, x_1^* > - \max_{x*\in B} < x, x^* >$ attain a maximum value of 0 at $x = 0$. This establishes (3) and (4).

Since

$$\min_{x\in A} f(x) = \max_{x_*\in B\cap X^*} g^*(x^*) = \max_{x*\in B} \min_{x\in A} < x, x* >$$

the theorem is proved.

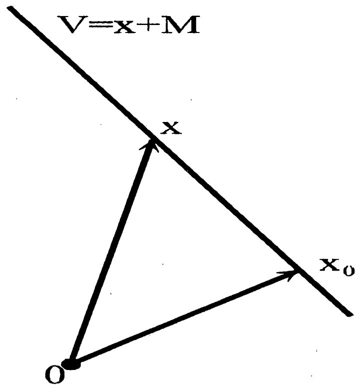
□

In the previous section we considered in some detail the problem of approximating an arbitrary vector in a Hilbert space by a vector in a given finite-dimensional subspace. The projection theorem led to the normal equations which could be solved for the best approximation. A major assumption in such problems was the finite dimensionality of the subspace from which the approximation was chosen. Finite dimensionality not only guarantees closure (and hence existence of a solution) but leads to a feasible computation procedure for obtaining the solution.

In many important and interesting practical problems the subspace in which the solution must lie is not finite dimensional. In such problems it is generally not possible to reduce the problem to a finite set of linear equations. However, there is an important class of such problems that can be reduced by the projection theorem to a finite set of linear equations similar to the normal equations. In this section we study these problems and their relation to the earlier approximation problem. We begin by pointing out a trivial modification of the projection theorem applicable to linear varieties.

Theorem: (Restatement of the Projection Theorem) Let M be a closed subspace of a Hilbert space H. Let x be a fixed element in H and let V be the linear variety $x + M$. Then there is a unique vector x_0 in V of minimum norm. Furthermore, x_0 is orthogonal to M.

Proof: This is proved by translating V by $-x$ so that it become a closed subspace and then applying the projection theorem. This is shown below.



Minimum norm to a linear variety

At this point we should note that the minimum norm solution x_0 is not orthogonal to the linear variety V but to the subspace M from which V is obtained.

A special kind of linear variety is of particular interest in optimization because it always leads to a finite-dimensional problems. This is the n-dimensionallinear varietyconsisting of points of the form $x + \sum_{i=1}^{n} a_i x_i$ where $\{x_1, x_2, x_3, \ldots, x_n\}$ is a linearly independent set in H, and x is a fixed vector in H. Problems which seek minimum norm vectors in an n-dimensional variety can be reduced to the solution of n - dimensional set of normal equations.

Chapter 4

OPTIMIZATION IN THE BANACH SPACE

Mathematicians take great care to formulate problems arising in applications as equivalent problems in Banach spaces rather than problems in other incomplete norm spaces. The principal advantage of a Banach space in optimization problems is that when seeking an optimal vector maximizing a given objective, it is possible to construct a sequence of vectors with each member superior to the preceding member, the desired optimal vector is then the limit of the sequence. In order that the scheme be effective and complete, the limit must be in the space, which is always true since a Banach space is complete.

In a Hilbert space, we can introduce the notion of orthogonal coordinates through an orthogonal base, and these coordinates are the values of bounded linear functionals defined by vectors of the base. The projection theorem which we used extensively to study optimization in the Hilbert space can be extended to a Banach space by the Hahn-Banach theorem. The Hahn-Banach Theorem, the most important theorem for the study of optimization in linear spaces, can be stated in several equivalent ways each having its own particular conceptual advantage. One of the forms called the 'extension form' serves as an appropriate generalization of the projection theorem from Hilbert space to normed spaces and thus provides a mean of generalizing many of the results of minimum norm problems. In a nutshell, this version extends the projection theorem to optimization problems having nonquadratic objectives. In this manner, the simpler geometric interpretation is preserved for the more complex problems. Another version, not discussed in this project, states in simpler form that given a sphere and a point not on the sphere there is a hyperplane separating the point and the sphere. This version together with the associated notions of hyperplanes and duality principles form a basis for many optimization problems in Banach Spaces.

Hahn-Banach Theorem

Before proving the extension version of Hahn-Banach theorem, we need the following definitions.

Definition: Let f be a linear functional defined on a subspace M of a vector space X. A linear functional F is said to be an **extension** of f if F is defined on a subspace N which property contains M, and if, on M, F is identical with f. In this case we say F is an extension of f from M to N.

In simple terms, the Hahn-Banach Theorem states that a bounded linear functional f defined on a subspace M of a normed linear space can be extended to a bounded linear functional F defined on the entire space and with norm equal to the norm of f on M, that is

$$\|F\| = \|f\|_M = \sup_{m \in M} \frac{|f(m)|}{\|m\|}$$

Definition: A real valued function p defined on a real vector space X is said to be a **sublinear functional** on X if

1. $p(x+y) \leq p(x) + p(y)$ for all $x, y \in X$
2. $p(\alpha x) = \alpha p(x)$ for $\alpha \geq 0$ and $x \in X$.

Theorem:(Hahn-Banach Theorem, Extension Version) Let X be a real linear normed space and p a continuous sublinear functional on X. Let f be a a linear functional defined on a space M of X satisfying $f(m) < p(m)$ for all $m \in M$. Then there is an extension F of f from M to X such that $F(x) \leq p(x)$ on X.

Proof: This theorem is true in any arbitrary normed linear space, but this version of the proof assums that X is separable. The general idea here is to extend f one dimension at a time and apply induction.

Suppose y is a vector in X but not in M. Consider all elements of the subspace $[M+y] = M \vee \{y\}$. Such an element x has a unique representation of the form $x = m + ay$, where $m \in M$ and a a real scalar. An extension g of f from M to $M \vee \{y\}$ has the form

$$g(x) = f(m) + ag(y)$$

and, hence, the extension is specified by prescribing the constant $g(y)$. We must show that this constant can be chosen so that $g(x) \leq p(x)$ on $M \vee \{y\}$.

For any two elements m_1 and m_2 in M, we have

$$f(m_1) + f(m_2) = f(m_1 + m_2) \leq p(m_1 + m_2) \leq p(m_1 - y) + p(m_2 + y)$$

or

$$f(m_1) - p(m_1 - y) \leq p(m_2 + y) - f(m_2)$$

and hence

$$\sup_{m \in M} [f(m_1) - p(m_1 - y)] \leq \inf_{m \in M} [p(m_2 + y) - f(m_2)]$$

Therefore there is a constant k such that

$$\sup_{m_1 \in M} [f(m_1) - p(m_1 - y)] \leq k \leq \inf_{m_2 \in M} [p(m_2 + y) - f(m_2)]$$

For any vector $x = m + ay \in M \vee \{y\}$, define $g(x) = f(m) + ak$. We need to show that $g(m + ay) \leq p(m + ay)$.

If a is positive, then

$$g(m+ay) = ak + f(m) = a\left[k + f\left(\tfrac{m}{a}\right)\right] \leq a\left[p\left(\tfrac{m}{a} + y\right) - f\left(\tfrac{m}{a}\right) + f\left(\tfrac{m}{a}\right)\right]$$
$$= ap\left(\tfrac{m}{a} + y\right) = p(m + ay)$$

If a is some negative number, say $a = -b < 0$, then

$$g(m - ay) = -bk + f(m) = b\left[-k + f\left(\tfrac{m}{b}\right)\right] \leq b\left[p\left(\tfrac{m}{b} - y\right) - f\left(\tfrac{m}{b}\right) + f\left(\tfrac{m}{b}\right)\right]$$
$$= bp\left(\tfrac{m}{b} - y\right) = p(m - by)$$

Thus $g(m + ay) \leq p(m+ay)$ for all a and g is an extension of f from M to $M \vee \{y\}$

Now let $\{x_1, x_2, \ldots, x_n, \ldots\}$ be a countable dense set in X. From this set of vectors select, one at a time , a subset of vectors, $\{y_1, y_2, \ldots, y_n, \ldots\}$ which is independent and independent of the subspace M. The set $\{y_1, y_2, \ldots, y_n, \ldots\}$ together with the subspace M generate a subspace S dense in X.

The functional f can be extended inductively to a functional g on the subspace S by extending f from M to $M \vee \{y_1\}$, then to $[[M \vee \{y_1\}] \vee \{y_2\}]$, and so on.

Finally the resulting g, which is continuous since p is, can be extended by continuity from the dense subspace S to the space X. Suppose $x \in X$, then there exists a sequence $\{x_n\}$ of vectors in S converging to x. Define $F(x) = \lim_{n\to\infty} g(x_n)$. F is obviously linear and $f(x) \leftarrow g(x_n) \leq p(x_n) \rightarrow p(x)$ so $F(x) \leq p(x)$ on X. □

Corollary 1: Let f be a bounded linear functional defined on a subspace M of a real normed vector space X. Then there is a bounded linear functional F defined on X which is an extension of F and which has norm equal equal to the norm of f on M.

Proof: Set $p(x) = \|f\|_M \|x\|$ in the Hahn-Banach Theorem. Then p is a continuous sublinear functional dominating f.

□

Minimum Norm Problems In General Norm Spaces

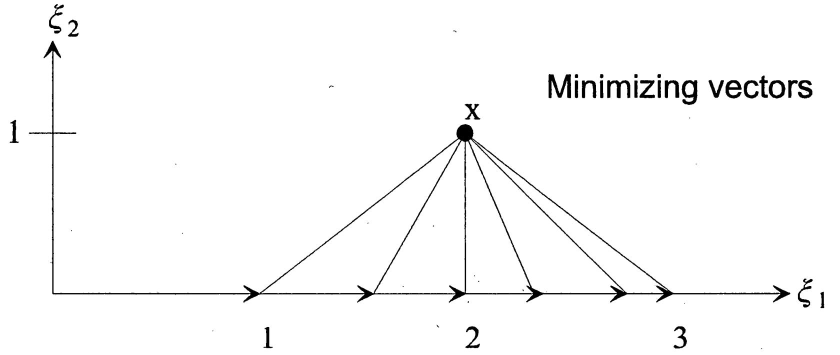
We end this section by considering the question of determining a vector in a subspace of a Normed space which best approximates a given vector x in the sense of minimum norm.

We recall that if M is a closed subspace in a Hilbert space, there is always a unique solution to the minimum norm problem and the solution satisfies the othorgonality condition. Forthermore the projection theorem leads to a linear equation for determining the unknown optimizing vector. In an arbitrary normed space, the optimizing vector, if it exist, may not be unique and the equations for the optimal vector will generally be nonlinear.

As an example that of the difficulties encountered in an arbitrary normed space, we consider a simple two dimensional minimum norm problem that does not have a single unique solution

Example: Let X be the space of pairs of real numbers $x = (\xi_1, \xi_2)$ with $\|x\| = \max_{i=1,2} |\xi_i|$. Then a convex set can have a minimum which is not unique.

Solution: Let M be a subspace of X consisting of all those vectors having their second components zero, that is $M = \{m = (a, 0)\}$ and consider the fixed point $x = (2, 1)$. The minimum distance from x to M is 1 since $\|x - m\| = 1 \; \forall a \in$ M. This situation is shown below.



The minimum norm theorem of this kind in general normed spaces, contain all the conclusions of the projection theorem except the uniqueness of the solution. When uniqueness holds it is fairly easy to show the set must be convex and closed. Uniqueness may be recovered if the normed space, satisfies the condition of **uniform convexity,** namely, if given any two elements x, y in a unit disc (that is each elements is of unit norm) such that

$$\|x{-}y\| \geq \epsilon > 0$$

there exist a δ greater than zero, depending only on ϵ, such that

$$\left\|\frac{x+y}{2}\right\| \leq 1 - \delta$$

Hille, E. and Phillips, R. have shown (5) that many of the properties of the minimum norm problems hold more generally in any Banach space with uniform convexity property. We note that any Hilbert space is trivially uniformly convex.

Furthermore, the solution to the minimum norm problem introduces a duality principle stating the equivalent of two extremization problems: one formulated in a normed space and the other in its dual. Often the transition from one problem to its dual, results in significant simplification or enhances physical and mathematical insight. Some infinite-dimensional problems can be converted to equivalent finite-dimensional problems by consideration of the dual problem. In a dual space there are two equivalent version of optimization problems. One in X called the primal problem and the other in X^* called the dual problem. The problems are related through both the optimal values of their respective objective functionals and an alignment condition on their solution vectors. Since in many spaces alignment can be explicitly characterized, the solution of either problem often lead directly to the solution of the other. Duality relations such as this are therefore often of extreme practical as well as theoretical significance in optimization problems. This is simply due to the fact that the Hahn-Banach theorem establishes the existence of certain linear functionals rather than vectors and establishes the general rule, that minimum norm problems must be formulated in a dual space if a solution existence is to be guaranteed.

BIBLIOGRAPHY

1. Balakrishnan, A. V., Introduction to Optimization Theory in a Hilbert Space, Springer, Berlin, 1971.
2. Balakrishnan, A. V., "Optimal Control problems in Hilbert Spaces." J.SIAM Control, Ser. A. 3(1) 152-185, 1965.
3. Berberian, S. K., Introduction to Hilbert Space, Oxford University Press, 1961.
4 Day, M. M., Normed Linear Spaces, Springer, Berlin, 1978.
5. Hille, E. and Phillips R., Functional Analysis and Semigroups, AMS Colloquium, Vol 31, 1957.
6. Kantorovich, L. V. and G. P. Akilov, Functional Analysis in Normed Spaces, Pergamon Press, Macmillan, New York, 1964.
7. Reed, M and Simon, Barry, Functional Analysis 1, Academic Press, New York, 1880.
8. Simmons, G., Introduction to Topology and Modern Analysis, Academic Press, 1776.