Chapter 1

Preliminaries

The purpose of this chapter is to provide some basic background information.

- Linear Space
- Hilbert Space
- Basic Principles
Linear Space

The notion of linear space provides simple, intuitive, and geometric interpretations on many complex mathematical entities.

• A linear space is an algebraic entity where every of its elements can be expressed algebraically as a linear combination of other elements.
  ◦ The simplest linear space is a vector space.
  ◦ A linear space can also be considered as a “flat” geometric entity in the sense of a “hyperplane” in the ambient space.

• If a linear space is equipped with a notion of sizes, called “norm”, of its elements, then
  ◦ We may measure the “distance” between any two elements.
  ◦ We may introduce the idea of “convergence”.
  ◦ Whether the limit points belong to a linear space is quite a significant issue.

• If a linear space is equipped with a notion of “inner product”, then
  ◦ The inner product can be used to induce a norm.
  ◦ The inner product can be used to determine whether two elements in the linear space are “perpendicular” or not.
  ◦ The notion of “projection” is critically important in applications.
A vector space \( V \) over a field \( \mathbb{F} \) is a set of elements called vectors together with two operations,

\[
\begin{align*}
+ : & \quad V \times V \to V \quad \text{via} \quad (x, y) \mapsto x + y, \\
\cdot : & \quad \mathbb{F} \times V \to V \quad \text{via} \quad (\alpha, x) \mapsto \alpha x,
\end{align*}
\]

that satisfy the following properties:

1. \( x + y = y + x \).
2. \( (x + y) + z = x + (y + z) \).
3. There is a null vector \( 0 \in V \) such that \( x + 0 = x \) for all \( x \in V \).
4. \( \alpha(x + y) = \alpha x + \alpha y \).
5. \( (\alpha + \beta)x = \alpha x + \beta x \).
6. \( (\alpha \beta)x = \alpha (\beta x) \).
7. \( 0x = 0, \ 1x = x \).

A nonempty subset \( S \) of a vector space \( V \) is called a subspace of \( V \) if every vector of the form \( \alpha x + \beta y \in S \) whenever \( x, y \in S \).
• Some examples:
  ◦ The set
    \[ \mathcal{P}_n := \{a_n x^n + \ldots + a_1 x + a_0 | a_n \in \mathbb{R}\} \]
    of all polynomials over \( \mathbb{R} \) with degree less than or equal to \( n \) is a vector space.
  ◦ The set
    \[ \mathcal{C}^1(a, b) := \{ f : (a, b) \to \mathbb{R} | f' \text{ is continuous over } (a, b) \} \]
    of continuously differentiable functions is a vector space.
  ◦ The set
    \[ \mathcal{T}_n := \{ T \in \mathbb{R}^{n \times n} | t_{ij} = r_{|i-j|+1}, \text{ where } r_1, \ldots, r_n \in \mathbb{R} \text{ arbitrary} \} \]
    of \( n \)-dimensional Toeplitz matrices is a vector space.
  ◦ The set
    \[ \mathcal{L}_p[a, b] := \{ f : [a, b] \to \mathbb{C} | \int_a^b |f(x)|^p dx < \infty \} \]
    of \( \mathcal{L}_p \) integrable functions is a vector space.
  ◦ The set
    \[ c_0 := \{ \{\xi_k\}_{k=1}^{\infty} | \lim_{k \to \infty} \xi_k = 0 \} \]
    is a vector space.
More Definitions

- A set $V$ is called a *linear variety* of a subspace $M$ if
  \[ V = x_0 + M = \{x_0 + m | m \in M\}. \]

- A set $C$ is said to be a *cone with vertex at the origin* if
  \[ \alpha x \in C, \text{ whenever } x \in C \text{ and } \alpha > 0. \]
  - The set of all nonnegative matrices is a cone.
  - The set of all nonnegative continuous functions is a convex cone.

- A vector $x$ is said to be *linearly independent* of a set $S$ if $x$ cannot be written as a linear combination of vectors from $S$.
  - The vectors $x_1, \ldots, x_n$ are mutually linearly independent if and only if
    \[ \sum_{i=1}^{n} \alpha_i x_i = 0 \implies \alpha_i = 0, \quad i = 1, \ldots, n. \]
  - If $x_1, \ldots, x_n$ are linearly independent, and $\sum_{i=1}^{n} \alpha_i x_i = \sum_{i=1}^{n} \beta_i x_i$, then $\alpha_i = \beta_i$ for all $i = 1, \ldots, n$.

- A finite set \{x_1, \ldots, x_n\} of linearly independent vectors is said to be a *basis* of the space $V$ if and only if
  \[ V = \text{span}\{x_1, \ldots, x_n\} := \{\sum_{i=1}^{n} \alpha_i x_i | \alpha_i \in \mathbb{F}\}. \]
Convexity

- A set $K$ is said to be convex if
  $$\alpha x + (1 - \alpha)y \in K, \text{ whenever } x, y \in K \text{ and } 0 \leq \alpha \leq 1.$$  
  - The sum of two convex sets is convex.
  - The intersection of convex sets is convex.

- Given a set $S$, the convex hull of $S$, denoted by $co(S)$, is defined to be the smallest convex set containing $S$.
  - Prove that $co(S)$ is the collection of all convex combinations from $S$, i.e.,
    $$co(S) = \{ \sum_{i=1}^{n} \alpha_i x_i | x_i \in S, \alpha_i \geq 0, \sum_{i=1}^{n} \alpha_i = 1, \text{ and } n \text{ is arbitrary but finite} \}.$$  
  - What is the convex hull of all $n$-dimensional orthogonal matrices?
  - What is the convex hull of all $n$-dimensional doubly stochastic matrices? (Birkhoff’s theorem)

- A real-valued function $\phi : V \to \mathbb{R}$ is said to be a convex function if
  $$\phi(\lambda x + (1 - \lambda)y) \leq \lambda \phi(x) + (1 - \lambda)y$$  
  for each $x, y \in V$ and $0 \leq \lambda \leq 1$. 
Normed Vector Space

- A vector space $V$ is said to be a *normed space* if there is a map
  \[ \| \cdot \| : V \to \mathbb{R} \]
  such that
  1. $\| x \| > 0$ if $x \neq 0$.
  2. $\| x + y \| \leq \| x \| + \| y \|$ for all $x$ and $y$ in $V$.
  3. $\| \alpha x \| = |\alpha| \| x \|$ for all $x \in V$ and $\alpha \in \mathbb{F}$.

- Some examples:
  - Over $\mathbb{R}^n$, the function
    \[ \| x \|_p := \left( \sum_{i=1}^{n} |x_i|^p \right)^{1/p} \]
    is a norm, if $p \geq 1$.
    - Identify the unit balls of $\mathbb{R}^n$ under different vector norms.
    - Note that unit balls are necessarily convex.
    - How does the notion of unit balls affects approximations?
Over $\mathbb{R}^{n \times n}$, the function
\[ \|M\|_p := \sup_{x \neq 0} \frac{\|Mx\|_p}{\|x\|_p} \]
is called an induced (matrix) norm.

- What is the geometric meaning of an induced matrix norm?
- What is the analog of the SVD if other norms are used in the variation formulation? Any usage?
- Why is the usual Frobenius norm of a matrix not an induced norm?

Over $C^1[a, b]$, the function
\[ \|f\| := \max_{a \leq x \leq b} |f(x)| + \max_{a \leq x \leq b} |f'(x)| \]
is a norm. (Note that the maximum exists over the closed interval $[a, b]$.)

- Define the total variation of a function $f$ over $[a, b]$ by
\[ T^b_a(f) := \sup_{a = x_0 < \ldots < x_n = b} \sum_{i=1}^{n} |f(x_i) - f(x_{i-1})| \]

- If $T^b_a(f) < \infty$, we say that $f$ is of bounded variation.
- The set
\[ BV[a, b] := \{ f : [a, b] \to \mathbb{R} | T^b_a(f) < \infty \} \]
is a normed vector space with the norm defined by
\[ \|f\| := f(a) + T^b_a(f). \]

- Give an example of a continuous function which is not of bounded variation.
A sequence of vector \( \{x_n\} \) in a normed vector space \( V \) is said to converge to a vector \( x \), denoted by \( x_n \to x \) if

\[
\|x_n - x\| \to 0 \quad \text{as } n \to \infty.
\]

- If a sequence converges, then its limit is unique.

- A sequence \( \{x_n\} \) in a normed vector space \( V \) is said to be a Cauchy sequence if to every \( \varepsilon > 0 \), there exists an integer \( N \) such that \( \|x_n - x_m\| < \varepsilon \) whenever \( m, n > N \).

- Every convergent sequence is a Cauchy sequence.
- A Cauchy sequence may not be convergent. (*Give an example!*)

- A space \( V \) in which every Cauchy sequence converges to a limit in \( V \) is said to be complete.
  - A complete normed vector space is called a Banach space.
Some examples:

- The space $C[a, b]$ with the norm

$$
\|f\| := \int_a^b |f(x)| \, dx
$$

is incomplete.

- The sequence

$$
f_n(x) := \begin{cases} 
0, & \text{for } 0 \leq x \leq \frac{1}{2} - \frac{1}{n}, \\
0x - \frac{n}{2} + 1, & \text{for } \frac{1}{2} - \frac{1}{n} \leq x \leq \frac{1}{2}, \\
1 & \text{for } x \geq \frac{1}{2}.
\end{cases}
$$

is Cauchy since $\|f_n - f_m\| = \frac{1}{2n} - \frac{1}{2m}$, but its limit point is not continuous.

- The same space $C[a, b]$ with the norm

$$
\|f\| := \max_{a \leq x \leq b} |f(x)|,
$$

the space $L^p[a, b]$ with $p \geq 1$, and any finite-dimensional normed vector space are complete.

- Any normed vector space $V$ is isometrically isomorphic to a dense subset of a Banach space $\tilde{V}$.
  
  - $X$ and $Y$ are isomorphic if there exist a one-to-one mapping $T$ of $X$ onto $Y$ such that

$$
T(a_1 x_1 + a_2 x_2) = a_1 T(x_1) + a_2 T(x_2).
$$

- An isomorphism $T$ is isometric if

$$
\|T(x)\| = \|x\|.
$$
A mapping $A$ from a vector $V$ into a vector space $W$ is said to be a **linear operator** if

$$A(\alpha x + \beta y) = \alpha A(x) + \beta A(y)$$

for all $x, y \in V$ and $\alpha, \beta \in \mathbb{F}$.

A linear operator $A$ is said to be **bounded** if

$$\|A\| := \sup_{x \in V; x \neq 0} \frac{\|Ax\|}{\|x\|} < \infty.$$ 

The above calculation can be simplified to $\|A\| = \sup_{\|x\|=1} \|Ax\|$.

A bounded linear operator is uniformly continuous.

If a linear operator is continuous at one point, it is bounded.

An example: Show that the differential operator

$$D(f) = f'$$

is linear and unbounded from $C^1(a, b)$ to $C(a, b)$ under some norms.

Given a normed vector space $X$ and a Banach space $Y$, the space

$$B(X, Y) := \{ A : X \rightarrow Y | A \text{ is linear and bounded} \}$$

is itself a Banach space with $\|A\|$ as the norm.
• The dual space $V^*$ of a normed vector space $V$ is defined to be

$$V^* := \{ f : V \to \mathbb{F} | f \text{ is linear and continuous} \}.$$ 

- The dual space $V^*$ of any normed vector space is automatically a Banach space.
  - Characterize the unit ball in $V^*$.

• (Hahn-Banach Theorem) This theorem plays a fundamental role in optimization theory.
  - Suppose that $f$ is a real-valued linear functional on a subspace $S$ and that $f(s) \leq p(s)$ for all $s \in S$.
  - Then there exists a linear functional $F : V \to \mathbb{R}$ such that $F(x) \leq p(x)$ for all $x \in V$ and $F(s) = f(s)$ for all $s \in S$.
    - What is the geometric meaning of the Hahn-Banach theorem in $\mathbb{R}^n$?
      - Pay attention to the minimum norm extension?
Hilbert Space

Hilbert spaces, equipped with their inner products, possess a wealth of geometric properties that generalize all of our understanding about the classical Euclidean space $\mathbb{R}^n$.

- A concept of orthogonality analogous to the “dot product” in $\mathbb{R}^3$ can be developed based on the “belief” of inner product. We use the algebraic expression of inner product to probe the inner structure of an abstract space.

- Orthogonality naturally leads to the notion of projection, which then leads to the idea of minimum distance. Though the minimum norm in a general normed vector space has a similar idea, the shortest distance by projection has a particularly appealing geometric intuition.
Inner Product

- Given a vector space $X$ over a field $\mathbb{F}$ (either $\mathbb{C}$ or $\mathbb{R}$), we say that it is equipped with an *inner product*

  $$\langle \cdot, \cdot \rangle : X \times X \rightarrow \mathbb{F}$$

if the inner product satisfies the following axioms:

1. $\langle x, y \rangle = \overline{\langle y, x \rangle}$.
2. $\langle x + y, z \rangle = \langle x, z \rangle + \langle y, z \rangle$.
3. $\langle \alpha x, y \rangle = \alpha \langle x, y \rangle$.
4. $\langle x, x \rangle \geq 0$ and $\langle x, x \rangle = 0$ if and only if $x = 0$.

- The quantity

  $$\|x\| := \langle x, x \rangle$$

  naturally defines a vector norm on $X$.

- Two vectors $x$ and $y$ are said to be *orthogonal*, denoted by $x \perp y$, if $\langle x, y \rangle = 0$. 
Some basic facts:

- (Cauchy-Schwarz Inequality) For all \(x, y\) in an inner product space,

\[
\langle x, y \rangle \leq \|x\| \|y\|.
\]

Equality holds if and only if \(x = \alpha y\) or \(y = 0\).

- (Parallelogram Law) With the induced inner product norm,

\[
\|x + y\|^2 + \|x - y\|^2 = 2\|x\|^2 + 2\|y\|^2.
\]

- (Continuity) Suppose that \(x_n \to x\) and \(y_n \to y\). Then \(\langle x_n, y_n \rangle \to \langle x, y \rangle\).

- Note that these geometric properties follow from the algebraic definition of inner product.

A complete inner product space is called a Hilbert space.
Orthogonal Complements

- Given any subset $S$ in an inner product space $X$, the set
  \[ S^\perp := \{ y \in X | y \perp x \text{ for every } x \in S \} \]
  is called the **orthogonal complement** of $S$.
  - $S^\perp$ is a closed subspace.
  - $S \subseteq (S^\perp)^\perp$.
  - $(S^\perp)^\perp$ is the smallest closed subspace containing $S$.

- We say that $X$ is the **direct sum** of two subspaces $M$ and $N$, denoted by $M = M \bigoplus N$, if every $x \in X$ has a unique representation of the form $x = m + n$ where $m \in M$ and $n \in N$.
  - The Classical Projection Theorem: Let $M$ be a closed subspace of a Hilbert space $H$. Corresponding to every $x \in H$,
    - There exists a unique $m_0 \in M$ such that $\| x - m_0 \| \leq \| x - m \|$ for all $m \in M$.
    - A necessary and sufficient condition that $m_0$ be the unique best approximation of $x$ in $M$ is that $x - m_0 \perp M$.
  - If $M$ is a closed linear subspace of a Hilbert space $X$, then $X = M \bigoplus M^\perp$.

- Given a vector $x_0 \in X$ and a closed subspace $M$, the unique vector $m_0 \in M$ such that $x_0 - m_0 \in M^\perp$ is called the **orthogonal projection** of $x_0$ onto $M$. 
Gram-Schmidt Orthogonalization Process

- Any given sequence (finite or infinite) \( \{x_n\} \) of linearly independent vectors in an inner product space \( X \) can be orthogonalized in the following sense.
  - There exists an orthonormal sequence \( \{z_n\} \) in \( X \) such that for each finite positive integer \( \ell \),
    \[
    \text{span}\{x_1, \ldots, x_\ell\} = \text{span}\{z_1, \ldots, z_\ell\}.
    \]
  - Indeed, the sequence \( \{z_n\} \) can be generated via the Gram-Schmidt process.
    \[
    z_1 := \frac{x_1}{\|x_1\|},
    \]
    \[
    w_n := x_n - \sum_{i=1}^{n-1} (x_i, z_i) z_i, \quad n > 1,
    \]
    \[
    z_n := \frac{w_n}{\|w_n\|}, \quad n > 1.
    \]
  - Show that in finite dimensional space, the Gram-Schmidt process is equivalent to that any full column rank matrix \( A \in \mathbb{R}^{n \times m} \) can be decomposed as
    \[
    A = QR
    \]
    where \( Q \in \mathbb{R}^{n \times m} \), \( Q^T Q = I_m \), and \( R \in \mathbb{R}^{m \times m} \), \( R \) is upper triangular.
• An example: Applying the Gram-Schmidt procedure with respect to a given inner product in the space $C^1[a, b]$ on the sequence of polynomials $\{1, x^1, x^2, \ldots\}$.

  ◦ Six popular orthogonal polynomial families:
    ▶ Gegenbuer: $[-1, 1]$, $\omega(x) = (1 - x^2)^{a - \frac{1}{2}}$, $a$ any irrational number of rational $\geq -\frac{1}{2}$.
    ▶ Hermite: $(-\infty, \infty)$, $\omega(x) = e^{-x^2}$.
    ▶ Laguerre: $[0, \infty)$, $\omega(x) = e^{-x}x^a$, $a$ any irrational or rational $\geq -1$.
    ▶ Legendre: Needed in Gaussian quadrature.
    ▶ Jacobi: $[-1, 1]$, $\omega(x) = (1 - x)^a(1 + x)^b$, $a$ any irrational or rational $\geq -1$.
    ▶ Chebyshev: $(-1, 1)$, $\omega(x) = (1 - x^2)^{\pm \frac{1}{2}}$.

  ◦ Show that the resulting orthonormal polynomials $\{z_n\}$ satisfy a recursive relationship of the form

\[
z_n(x) = (a_n x + b_n)z_{n-1}(x) - c_n z_{n-2}(x), \quad n = 2, 3, \ldots
\]

where the coefficients $a_n, b_n, c_n$ can be explicitly determined.

  ◦ Show that the zeros of the orthonormal polynomials are real, simple, and located in the interior of $[a, b]$.
One of the most fundamental optimization question is as follows:

- Let \( x_0 \) represent a target vector in a Hilbert space \( X \).
  - The target vector could mean a true solution that is hard to get in the abstract space \( X \).
- Let \( M \) denote a subspace of \( X \).
  - Consider \( M \) as the set of all computable, constructible, or reachable (by some reasonable means) vectors in \( X \).
  - \( M \) shall be called a feasible set.
- Want to solve the optimization problem

\[
\min_{m \in M} \|x_0 - m\|. 
\]
Preliminaries

Projection Theorem

- A necessary and sufficient condition that \( m_0 \in M \) solves the above optimization problem is that
  \[ x_0 - m_0 \perp M. \]
  - The minimizer \( m_0 \) is unique, if it exists.
  - The existence is guaranteed only if \( M \) is closed.

- How to find this minimizer?
  - Assume that \( M \) is of finite dimension and has a basis
    \[ M = \text{span}\{ y_1, \ldots, y_n \}. \]
    Write
    \[ m_0 = \sum_{i=1}^{n} \alpha_i y_i. \]
    Then \( \alpha_1, \ldots, \alpha_n \) satisfy the normal equation
    \[ \langle x_0 - \sum_{i=1}^{n} \alpha_i y_i, y_j \rangle = 0, \quad j = 1, \ldots, n. \]
In matrix form, the linear system can be written as
\[
\begin{pmatrix}
\langle y_1, y_1 \rangle & \langle y_2, y_1 \rangle & \cdots & \langle y_n, y_1 \rangle \\
\langle y_1, y_2 \rangle & \langle y_2, y_2 \rangle & \cdots & \langle y_n, y_2 \rangle \\
\vdots & \vdots & \ddots & \vdots \\
\langle y_1, y_n \rangle & \langle y_2, y_n \rangle & \cdots & \langle y_n, y_n \rangle \\
\end{pmatrix}
\begin{pmatrix}
\alpha_1 \\
\alpha_2 \\
\vdots \\
\alpha_n \\
\end{pmatrix}
=
\begin{pmatrix}
\langle x_0, y_1 \rangle \\
\langle x_0, y_2 \rangle \\
\vdots \\
\langle x_0, y_n \rangle \\
\end{pmatrix}.
\]

If the basis \(\{y_1, \ldots, y_n\}\) are orthonormal, then trivially
\[
\alpha_i = \langle x_0, y_i \rangle.
\]

What can be done if \(M\) is not of finite dimension?

If \(M\) is of codimension \(n\), i.e., if the orthogonal complement of \(M\) has dimension \(n\), then a dual approximation problem can be formulated.
Fourier Series

- Let \( \{z_n\} \) be an orthonormal sequence in a Hilbert space \( X \).
  - The convergence of an infinite series in \( X \) can be identified as a square-summable sequence in \( \mathbb{R} \) in the following way:
    \[
    \sum_{i=1}^{\infty} \xi_i z_i \longrightarrow x \quad \text{if and only if} \quad \sum_{i=1}^{\infty} |\xi_i|^2 < \infty.
    \]
  - \( \xi_i = \langle x, z_i \rangle \).

- (Bessel’s Inequality) Given \( x \in X \), then
  \[
  \sum_{i=1}^{\infty} |\langle x, z_i \rangle|^2 \leq \|x\|^2.
  \]

- The convergent series
  \[
  \mathcal{F}(x) := \sum_{i=1}^{\infty} \langle x, z_i \rangle z_i
  \]
  guaranteed by Bessel’s inequality is called the *Fourier series* of \( x \) in \( X \).
  - \( \mathcal{F}(x) \) belongs to the closure of \( \text{span}\{z_n\} \).
  - \( x - \mathcal{F}(x) \perp \text{span}\{z_n\} \) (and hence its closure).
  - When will an orthonormal sequence \( \{z_n\} \) generate a Hilbert space?
• An orthonormal sequence \( \{z_n\} \) in a Hilbert space \( X \) is said to be complete if the closure of \( \text{span}\{z_n\} \) is \( X \) itself.

○ An orthonormal sequence \( \{z_n\} \) in a Hilbert space is complete if and only if the only vector orthogonal to each \( z_n \) is the zero vector.

○ The subspace of polynomials is dense in \( \mathcal{L}_2[-1, 1] \).

○ The Legendre polynomials, i.e., the orthonormal sequence

\[
z_n(x) = \sqrt{\frac{2n + 1}{2}} \frac{(-1)^n}{2^n n!} \frac{d^n}{dx^n} (1 - x^2)^n
\]

is a complete in \( \mathcal{L}_2[-1, 1] \).
Minimum Norm Problems

The minimization problem described in the proceeding section can appear in different forms, including

- The dual approximation problem.
- The minimum energy control problem.
- The shortest distance to a convex set.
Dual Problem of Best Approximation

- Given \( \{y_1, \ldots, y_n\} \) in a Hilbert space, define the feasible set
  \[
  F := \{ x \in H | \langle x, y_i \rangle = c_i \}.
  \]
  Want to solve the minimum norm problem
  \[
  \min_{x \in F} \|x\|.
  \]
- Define
  \[
  M := \text{span}\{y_1, \ldots, y_n\}.
  \]
  Then
  \[
  F = c + M^\perp,
  \]
  for some suitable \( c \in F \).
  - Analogous to the notion that a general solution is the sum of a particular solution \( c \) and a general homogeneous solution \( M^\perp \).
  - We say that \( F \) is a linear variety with codimension \( n \).
    - Note that \( F \) (or \( M^\perp \)) itself could be of infinite dimension.
• Using the Projection Theorem, we know that the solution exists and is unique. Indeed, the optimal solution \( x^* \)

\[
x^* \in M^\perp = M.
\]

\diamond \; x^* = \sum_{i=1}^{n} \beta_i y_i.

\diamond \; The coefficients \( \beta_i \) are determined from the linear system

\[
\begin{pmatrix}
\langle y_1, y_1 \rangle & \langle y_2, y_1 \rangle & \cdots & \langle y_n, y_1 \rangle \\
\langle y_1, y_2 \rangle & \cdots & \cdots & \cdots \\
\vdots & \cdots & \cdots & \cdots \\
\langle y_1, y_n \rangle & \langle y_n, y_1 \rangle & \cdots & \langle y_n, y_n \rangle \\
\end{pmatrix}
\begin{pmatrix}
\beta_1 \\
\beta_2 \\
\vdots \\
\beta_n \\
\end{pmatrix}
= \begin{pmatrix}
c_1 \\
c_2 \\
\vdots \\
c_n \\
\end{pmatrix}.
\]

\diamond \; What is the geometric meaning of the dual problem?
A Control Problem

- Consider the problem of minimizing

\[ J(x, u) := \int_0^T (x^2(t) + u^2(t))dt \]

subject to

\[ \dot{x}(t) = u(t), \quad x(0) = \text{given}. \]

- The objective function represents a compromise between a desire to have \( x(t) \) small while simultaneously conserving control energy.

- Equip the product space \( H = L^2[0, T] \times L^2[0, T] \) with the inner product

\[ \langle (x_1, u_1), (x_2, u_2) \rangle = \int_0^T [x_1(t)x_2(t) + u_1(t)u_2(t)]dt. \]

- Let \( V \) be the linear variety

\[ V := \{ (x, u) \in H | x(t) = x(0) + \int_0^t u(\tau)d\tau \}. \]

- The control problem is equivalent to finding the element in \( V \) with minimal norm.

  - It can be shown that \( V \) is closed in \( H \).
  - The Projection Theorem is hence applicable. The solution exists and is unique.
  - How to do the computation?
Minimum Distance to a Convex Set

- The concept of linear varieties can be generalized to convex sets.
- Given a vector $\mathbf{x}$ and a closed convex subset $K$ in a Hilbert space $H$,
  - There is a unique $k_0 \in K$ such that
    \[ \| \mathbf{x} - k_0 \| \leq \| \mathbf{x} - k \| \]
    for all $k \in K$.
  - A necessary and sufficient condition for $k_0$ being the global minimizer is that
    \[ \langle \mathbf{x} - k_0, k - k_0 \rangle \leq 0 \]
    for all $k \in K$. 
Linear Complementary Problem

- Given \( \{y_1, \ldots, y_n\} \subset H \), and \( x \in H \), want to minimize

\[
\|x - \sum_{i=1}^{n} \alpha_i y_i \|
\]

subject to \( \alpha_i \geq 0 \) for all \( i \).
- The feasible set forms a cone.
  - There exists a unique solution \( \hat{x} = \sum_i^n \alpha_i y_i \).
- Try to characterize the coefficients \( \alpha_i \).
  - Take \( k = \hat{x} + y_j \), then
    \[
    \langle x - \hat{x}, y_j \rangle \leq 0.
    \]
  - Take \( k = \hat{x} - \alpha_j y_j \), then
    \[
    \langle x - \hat{x}, -\alpha_j y_j \rangle \leq 0.
    \]
  - Together,
    \[
    \langle x - \hat{x}, y_i \rangle \leq 0, \text{ with equality if } \alpha_i > 0.
    \]
• Rewrite in matrix form,

\[
\begin{pmatrix}
\langle y_1, y_1 \rangle & \langle y_2, y_1 \rangle & \cdots & \langle y_n, y_1 \rangle \\
\langle y_1, y_2 \rangle & \ddots & \vdots & \langle y_1, y_n \rangle \\
\vdots & \ddots & \ddots & \vdots \\
\langle y_1, y_n \rangle & \cdots & \langle y_n, y_n \rangle & \langle y_1, y_1 \rangle
\end{pmatrix}
\begin{pmatrix}
\alpha_1 \\
\alpha_2 \\
\vdots \\
\alpha_n
\end{pmatrix}
- 
\begin{pmatrix}
\langle x, y_1 \rangle \\
\langle x, y_2 \rangle \\
\vdots \\
\langle x, y_n \rangle
\end{pmatrix}
= 
\begin{pmatrix}
z_1 \\
z_2 \\
\vdots \\
z_n
\end{pmatrix}.
\]

\diamond \ z_i \geq 0 \text{ for all } i.

\diamond \ \alpha_i z_i = 0 \text{ for all } i.

• Check out CPnet for more discussions on complementary problems.
Quadratic Programming Problem

- Given a symmetric positive-definite matrix $Q \in \mathbb{R}^{n \times n}$, $A \in \mathbb{R}^{m \times n}$, $m < n$, and $b \in \mathbb{R}^m$, want to minimize
  
  $$x^T Q x,$$

  subject to $Ax = b$.

- The feasible set is a linear variety,

  $$Ax = b \iff x \in x_0 + \mathcal{N}(A).$$

  $\bigcirc \mathcal{N} = \mathcal{R}(A^T)^\perp.$

- This is a minimum norm problem.

  $\bigcirc \langle x, y \rangle = x^T Q y$ defines an (oblique) inner product.

- Write down a numerical algorithm for the solution.
Basic Principles in Optimization

According to David Luenberger, the theory of optimization can be summarized from a few “simple, intuitive, geometric relations.” These are

- The Projection Theorem.
  - The simplest form of this theorem is that the shortest line from a point to a plane is necessarily perpendicular to the plane.

- The Hahn-Banach Theorem.
  - The simplest form of this theorem is that a sphere and a point not in the sphere can be “separated” by a hyperplane.

- Duality
  - The simplest way to describe the duality is that the shortest distance from a point to a convex set is equal to the maximum of the distance from the point to a hyperplane separating the point from the convex set.

- Differentials.
  - In $\mathbb{R}^n$, a functional is optimized only at places where the gradient vanishes.

Each of these notions can easily be understood in $\mathbb{R}^3$ and can be extended into infinite-dimensional space.