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.

Vector Space

- A *vector space* V over a field $\mathbb{F}$ is a set of elements called *vectors* together with *two* operations,

$$+: \quad V \times V \to V \qquad \text{via } (\mathbf{x}, \mathbf{y}) \rightarrowtail \mathbf{x} + \mathbf{y}, \tag{1}$$

$$.: \quad \mathbb{F} \times V \to V \qquad \text{via } (\alpha, \mathbf{x}) \rightarrowtail \alpha\mathbf{x}, \tag{2}$$

  that satisfy the following properties:

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

- A nonempty subset S of a vector space V is called a *subspace* of V if every vector of the form $\alpha\mathbf{x} + \beta\mathbf{y} \in S$ whenever $\mathbf{x}, \mathbf{y} \in S$.

- 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

$$V = \mathbf{x}_0 + M = \{\mathbf{x}_0 + \mathbf{m} | \mathbf{m} \in M\}.$$

- A set C is said to be a *cone with vertex at the origin* if

$$\alpha \mathbf{x} \in C, \quad \text{whenever } \mathbf{x} \in C \text{ and } \alpha > 0.$$

  - ⋄ The set of all nonnegative matrices is a cone.
  - ⋄ The set of all nonnegative continuous functions is a convex cone.

- A vector $\mathbf{x}$ is said to be *linearly independent* of a set S is $\mathbf{x}$ cannot be written as a linear combination of vectors from S.

  - ⋄ The vectors $\mathbf{x}_1, \ldots, \mathbf{x}_n$ are mutually linearly independent if and only if

$$\sum_{i=1}^{n} \alpha_i \mathbf{x}_i = 0 \Longrightarrow \alpha_i = 0, \quad i = 1, \ldots, n.$$

  - ⋄ If $\mathbf{x}_1, \ldots, \mathbf{x}_n$ are linearly independent, and $\sum_{i=1}^{n} \alpha_i \mathbf{x}_i = \sum_{i=1}^{n} \beta_i \mathbf{x}_i$, then $\alpha_i = \beta_i$ for all $i = 1, \ldots, n$.

- A finite set $\{\mathbf{x}_1, \ldots, \mathbf{x}_n\}$ of linearly independent vectors is said to be a *basis* of the space V if and only if

$$V = \operatorname{span}\{\mathbf{x}_1, \ldots, \mathbf{x}_n\} := \{\sum_{i=1}^{n} \alpha_i \mathbf{x}_i | \alpha_i \in \mathbb{F}\}.$$

Convexity

- A set K is said to be *convex* if

$$\alpha\mathbf{x} + (1-\alpha)\mathbf{y} \in K, \quad \text{whenever } \mathbf{x}, \mathbf{y} \in K \text{ and } 0 \leq \alpha \leq 1.$$

  - ◇ 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.,

$$co(S) = \{\sum_{i=1}^{n} \alpha_i \mathbf{x}_n | \mathbf{x}_i \in S, \alpha_i \geq 0, \sum_{i=1}^{n} \alpha_i = 1, \text{ and } n \text{ is arbitrary but finite}\}.$$

  - ◇ 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

$$\phi(\lambda\mathbf{x} + (1-\lambda)\mathbf{y}) \leq \lambda\phi(\mathbf{x}) + (1-\lambda)\mathbf{y}$$

  for each $\mathbf{x}, \mathbf{y} \in V$ 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. $\|\mathbf{x}\| > 0$ if $\mathbf{x} \neq \mathbf{0}$.
  2. $\|\mathbf{x} + \mathbf{y}\| \leq \|\mathbf{x}\| + \|\mathbf{y}\|$ for all $\mathbf{x}$ and $\mathbf{y}$ in V.
  3. $\|\alpha \mathbf{x}\| = |\alpha| \|\mathbf{x}\|$ for all $\mathbf{x} \in V$ and $\alpha \in \mathbb{F}$.

- Some examples:
  - ⋄ Over $\mathbb{R}^n$, the function

$$\|\mathbf{x}\|_p := \left(\sum_{i=1}^{n} |x_i|^p \right)^{1/p}$$

    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

$$\|M\|_p := \sup_{\mathbf{x}\neq\mathbf{0}} \frac{\|M\mathbf{x}\|_p}{\|\mathbf{x}\|_p}$$

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

$$\|f\| := \max_{a\le x\le b} |f(x)| + \max_{a\le x\le b} |f'(x)|$$

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

$$T_a^b(f) := \sup_{a=x_0<\ldots<x_n=b} \sum_{i=1}^{n} |f(x_i) - f(x_{i-1})|$$

  - If $T_a^b(f) < \infty$, we say that f is of *bounded variation*.
  - The set

$$BV[a,b] := \{f : [a,b] \to \mathbb{R} | T_a^b(f) < \infty\}$$

is a normed vector space with the norm defined by

$$\|f\| := f(a) + T_a^b(f).$$

  - Give an example of a continuous function which is not of bounded variation.

Convergence and Completeness

- A sequence of vector $\{\mathbf{x}_n\}$ in a normed vector space V is said to *converge* to a vector $\mathbf{x}$, denoted by $\mathbf{x}_n \to \mathbf{x}$ if
$$\|\mathbf{x}_n - \mathbf{x}\| \to 0 \quad \text{as } n \to \infty.$$
  - ◇ If a sequence converges, then its limit is unique.
- A sequence $\{\mathbf{x}_n\}$ in a normed vector space V is said to be a *Cauchy sequence* if to every $\epsilon > 0$, there exists and integer N such that $\|\mathbf{x}_n - \mathbf{x}_m\| < \epsilon$ 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

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

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 $\mathcal{L}_p[a,b]$ 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 $\hat{V}$.
    - ▷ X and Y are isomorphic if there exist a one-to-one mapping T of X onto Y such that

$$T(\alpha_1 \mathbf{x}_1 + \alpha_2 \mathbf{x}_2) = \alpha_1 T(\mathbf{x}_1) + \alpha_2 T(\mathbf{x}_2).$$

    - ▷ An isomorphism T is isometric if

$$\|T(\mathbf{x})\| = \|\mathbf{x}\|.$$

Banach Space

- A mapping A from a vector V into a vector space W is said to be a *linear operator* if

$$A(\alpha\mathbf{x} + \beta\mathbf{y}) = \alpha A(\mathbf{x}) + \beta A(\mathbf{y})$$

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

- A linear operator Λ is said to be *bounded* if

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

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

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

is itself a Banach space with $\|A\|$ as the norm.

- The *dual space* V^* of a normed vector space V is defined to be

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

  - 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.
  - Let $p : V \longrightarrow \mathbb{R}$ be a function satisfying $p(x+y) \leq p(\mathbf{x}) + p(\mathbf{y})$ and $p(\alpha\mathbf{x}) = \alpha p(\mathbf{x})$.
  - Suppose that f is a real-valued linear functional on a subspace S and that $f(\mathbf{s}) \leq p(\mathbf{s})$ for all $\mathbf{s} \in S$.
  - Then there exists a linear functional $F : V \longrightarrow \mathbb{R}$ such that $F(\mathbf{x}) \leq p(\mathbf{x})$ for all $\mathbf{x} \in V$ and $F(\mathbf{s}) = f(\mathbf{s})$ for all $\mathbf{s} \in S$.
    - 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*
$$\langle \cdot, \cdot \rangle : X \times X \to \mathbb{F}$$
if the inner product satisfies the following axioms:
  1. $\langle \mathbf{x}, \mathbf{y} \rangle = \overline{\langle \mathbf{y}, \mathbf{x} \rangle}$.
  2. $\langle \mathbf{x} + \mathbf{y}, \mathbf{z} \rangle = \langle \mathbf{x}, \mathbf{z} \rangle + \langle \mathbf{y}, \mathbf{z} \rangle$.
  3. $\langle \alpha \mathbf{x}, \mathbf{y} \rangle = \alpha \langle \mathbf{x}, \mathbf{y} \rangle$.
  4. $\langle \mathbf{x}, \mathbf{x} \rangle \geq 0$ and $\langle \mathbf{x}, \mathbf{x} \rangle = 0$ if and only if $\mathbf{x} = \mathbf{0}$.

  ◇ The quantity
  $$\|\mathbf{x}\| := \langle \mathbf{x}, \mathbf{x} \rangle$$
  naturally defines a vector norm on X.

  ◇ Two vectors $\mathbf{x}$ and $\mathbf{y}$ are said to be *orthogonal*, denoted by $\mathbf{x} \perp \mathbf{y}$, if $\langle \mathbf{x}, \mathbf{y} \rangle = 0$.

- Some basic facts:
  - ◇ (Cauchy-Schwarz Inequality) For all $\mathbf{x}, \mathbf{y}$ in an inner product space,

    $$\langle \mathbf{x}, \mathbf{y} \rangle \leq \|\mathbf{x}\| \|\mathbf{y}\|.$$

    Equality holds if and only if $\mathbf{x} = \alpha \mathbf{y}$ or $\mathbf{y} = \mathbf{0}$.
  - ◇ (Parallelogram Law) With the induced inner product norm,

    $$\|\mathbf{x} + \mathbf{y}\|^2 + \|\mathbf{x} - \mathbf{y}\|^2 = 2\|\mathbf{x}\|^2 + 2\|\mathbf{y}\|^2.$$

  - ◇ (Continuity) Suppose that $\mathbf{x}_n \to \mathbf{x}$ and $\mathbf{y}_n \to \mathbf{y}$. Then $\langle \mathbf{x}_n, \mathbf{y}_n \rangle \to \langle \mathbf{x}, \mathbf{y} \rangle$.
  - ◇ 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

$$S^{\perp} := \{\mathbf{y} \in X | \mathbf{y} \perp \mathbf{x} \text{ for every } \mathbf{x} \in S\}$$

  is called the *orthogonal complement* of S.

  - ◇ $S^{\perp}$ is a closed subspace.
  - ◇ $S \subset (S^{\perp})^{\perp}$.
  - ◇ $(S^{\perp})^{\perp}$ 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 $\mathbf{x} \in X$ has a unique representation of the form $\mathbf{x} = \mathbf{m} + \mathbf{n}$ where $\mathbf{m} \in M$ and $\mathbf{n} \in N$.

  - ◇ The Classical Projection Theorem: Let M be a closed subspace of a Hilbert space H. Corresponding to every $\mathbf{x} \in H$,
    - ▷ There exists a unique $\mathbf{m}_0 \in M$ such that $\|\mathbf{x} - \mathbf{m}_0\| \leq \|\mathbf{x} - \mathbf{m}\|$ for all $\mathbf{m} \in M$.
    - ▷ A necessary and sufficient condition that $\mathbf{m}_0$ be the unique best approximation of $\mathbf{x}$ in M is that $\mathbf{x} - \mathbf{m}_0 \perp M$.
  - ◇ If M is a closed linear subspace of a Hilbert space X, then $X = M \bigoplus M^{\perp}$.

- Given a vector $\mathbf{x}_0 \in X$ and a closed subspace M, the unique vector $\mathbf{m}_0 \in M$ such that $\mathbf{x}_0 - \mathbf{m}_0 \in M^{\perp}$ is called the *orthogonal projection* of $\mathbf{x}_0$ onto M.

Gram-Schmidt Orthogalization Process

- Any given sequence (finite or infinite) $\{\mathbf{x}_n\}$ of linearly independent vectors in an inner product space X can be orthogonalized in the following sense.
  - ◇ There exists an orthonormal sequence $\{\mathbf{z}_n\}$ in X such that for each finite positive integer ℓ,
    $$\text{span}\{\mathbf{x}_1, \ldots, \mathbf{x}_\ell\} = \text{span}\{\mathbf{z}_1, \ldots, \mathbf{z}_\ell\}.$$
  - ◇ Indeed, the sequence $\{\mathbf{z}_n\}$ can be generated via the Gram-Schmidt process.
    $$\begin{aligned} \mathbf{z}_1 &:= \frac{\mathbf{x}_1}{\|\mathbf{x}_1\|}, \\ \mathbf{w}_n &:= \mathbf{x}_n - \sum_{i=1}^{n-1} \langle \mathbf{x}_i, \mathbf{z}_i \rangle \mathbf{z}_i, \quad n > 1, \\ \mathbf{z}_n &:= \frac{\mathbf{w}_n}{\|\mathbf{w}_n\|}, \quad n > 1. \end{aligned}$$
  - ◇ 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 ≥ -1.
    - ▷ Legendre: Needed in Gaussian quadrature.
    - ▷ Jacobi: $[-1,1]$, $\omega(x) = (1-x)^a(1+x)^b$, a any irrational or ration ≥ -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 $[a,b]$.

Best Approximation

One of the most fundamental optimization question is as follows:

- Let $\mathbf{x}_0$ 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

$$\min_{\mathbf{m}\in M} \|\mathbf{x}_0 - \mathbf{m}\|.$$

Projection Theorem

- A necessary and sufficient condition that $\mathbf{m}_0 \in M$ solves the above optimization problem is that
$$\mathbf{x}_0 - \mathbf{m}_0 \perp M.$$
  - ◇ The minimizer $\mathbf{m}_0$ 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
$$M = \text{span}\{\mathbf{y}_1, \ldots, \mathbf{y}_n\}.$$
Write
$$\mathbf{m}_0 = \sum_{i=1}^{n} \alpha_i \mathbf{y}_i.$$
Then $\alpha_1, \ldots, \alpha_n$ satisfy the *normal equation*
$$\langle \mathbf{x}_0 - \sum_{i=1}^{n} \alpha_i \mathbf{y}_i, \mathbf{y}_j \rangle = 0, \quad j = 1, \ldots n.$$

◇ In matrix form, the linear system can be written as

$$\begin{pmatrix} \langle \mathbf{y}_1, \mathbf{y}_1 \rangle & \langle \mathbf{y}_2, \mathbf{y}_1 \rangle & \dots & \langle \mathbf{y}_n, \mathbf{y}_1 \rangle \\ \langle \mathbf{y}_1, \mathbf{y}_2 \rangle & & & \\ \vdots & & \ddots & \vdots \\ \langle \mathbf{y}_1, \mathbf{y}_n \rangle & & & \langle \mathbf{y}_n, \mathbf{y}_n \rangle \end{pmatrix} \begin{pmatrix} \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_n \end{pmatrix} = \begin{pmatrix} \langle \mathbf{x}_0, \mathbf{y}_1 \rangle \\ \langle \mathbf{x}_0, \mathbf{y}_2 \rangle \\ \vdots \\ \langle \mathbf{x}_0, \mathbf{y}_n \rangle \end{pmatrix}.$$

▷ If the basis $\{\mathbf{y}_1, \dots, \mathbf{y}_n\}$ are orthonormal, then trivially

$$\alpha_i = \langle \mathbf{x}_0, \mathbf{y}_i \rangle.$$

◇ 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 $\{\mathbf{z}_n\}$ 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:

$$\sum_{i=1}^{\infty} \xi_i \mathbf{z}_i \longrightarrow \mathbf{x} \quad \text{if and only if} \quad \sum_{i=1}^{\infty} |\xi_i|^2 < \infty.$$

  - ◇ $\xi_i = \langle \mathbf{x}, \mathbf{z}_i \rangle$.
- (Bessel's Inequality) Given $\mathbf{x} \in X$, then

$$\sum_{i=1}^{\infty} |\langle \mathbf{x}, \mathbf{z}_i \rangle|^2 \leq \|\mathbf{x}\|^2.$$

  - ◇ The convergent series

$$\mathcal{F}(\mathbf{x}) := \sum_{i=1}^{\infty} \langle \mathbf{x}, \mathbf{z}_i \rangle \mathbf{z}_i$$

    guaranteed by Bessle's inequality is called the *Fourier series* of $\mathbf{x}$ in X.
    - ▷ $\mathcal{F}(\mathbf{x})$ belongs to the closure of $\text{span}\{\mathbf{z}_n\}$.
    - ▷ $\mathbf{x} - \mathcal{F}(\mathbf{x}) \perp \text{span}\{\mathbf{z}_n\}$ (and hence its closure).
    - ▷ When will an orthonormal sequence $\{\mathbf{z}_n\}$ generate a Hilbert space?

- An orthonormal sequence $\{\mathbf{z}_n\}$ in a Hilbert space X is said to be *complete* if the closure of *span*$\{\mathbf{z}_n\}$ is X itself.
  - ◇ An orthonormal sequence $\{\mathbf{z}_n\}$ in a Hilbert space is complete if and only if the only vector orthogonal to each $\mathbf{z}_n$ is the zero vector.
  - ◇ The subspace of polynomials is dense in $\mathcal{L}_2[-1,1]$.
  - ◇ The Legendre polynomials, i.e., the orthonormal sequence

$$\mathbf{z}_n(x) = \sqrt{\frac{2n+1}{2}}\frac{(-1)^n}{2^n n!}\frac{d^n}{dx^n}(1-x^2)^n$$

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 $\{\mathbf{y}_1, \ldots, \mathbf{y}_n\}$ in a Hilbert space, define the feasible set

$$F := \{\mathbf{x} \in H | \langle \mathbf{x}, \mathbf{y}_i \rangle = c_i\}.$$

  Want to solve the minimum norm problem

$$\min_{\mathbf{x} \in F} \|\mathbf{x}\|.$$

- Define

$$M := \operatorname{span}\{\mathbf{y}_1, \ldots, \mathbf{y}_n\}.$$

  Then

$$F = \mathbf{c} + M^{\perp},$$

  for some suitable $\mathbf{c} \in F$.

  - ◇ Analogous to the notion that a general solution is the sum of a particular solution $\mathbf{c}$ 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 $\mathbf{x}^*$

$$\mathbf{x}^* \in M^{\perp\perp} = M.$$

  - ◇ $\mathbf{x}^* = \sum_{i=1}^{n} \beta_i \mathbf{y}_i$.
  - ◇ The coefficients β_i are determined from the linear system

$$\begin{pmatrix} \langle \mathbf{y}_1, \mathbf{y}_1 \rangle & \langle \mathbf{y}_2, \mathbf{y}_1 \rangle & \dots & \langle \mathbf{y}_n, \mathbf{y}_1 \rangle \\ \langle \mathbf{y}_1, \mathbf{y}_2 \rangle & & & \\ \vdots & & \ddots & \vdots \\ \langle \mathbf{y}_1, \mathbf{y}_n \rangle & & & \langle \mathbf{y}_n, \mathbf{y}_n \rangle \end{pmatrix} \begin{pmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_n \end{pmatrix} = \begin{pmatrix} c_1 \\ c_2 \\ \vdots \\ c_n \end{pmatrix}.$$

  - ◇ 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 comprise between a desire to have $x(t)$ small while simultaneously conserving control energy.

- Equip the product space $H = \mathcal{L}_2[0,T] \times \mathcal{L}_2[0,T]$ 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 $\mathbf{k}_0 \in K$ such that
  $$\|\mathbf{x} - \mathbf{k}_0\| \leq \|\mathbf{x} - \mathbf{k}\|$$
  for all $\mathbf{k} \in K$.
  - ◇ A necessary and sufficient condition for $\mathbf{k}_0$ being the global minimizer is that
  $$\langle \mathbf{x} - \mathbf{k}_0, \mathbf{k} - \mathbf{k}_0 \rangle \leq 0$$
  for all $\mathbf{k} \in K$.

Linear Complementary Problem

- Given $\{\mathbf{y}_1, \ldots, \mathbf{y}_n\} \subset H$, and $\mathbf{x} \in H$, want to minimize

$$\|\mathbf{x} - \sum_{i=1}^{n} \alpha_i \mathbf{y}_i\|,$$

subject to $\alpha_i \geq 0$ for all i.

- The feasible set forms a cone.
  - ⋄ There exists a unique solution $\hat{\mathbf{x}} = \sum_i^n \alpha_i \mathbf{y}_i$.
- Try to characterize the coefficients α_i.
  - ⋄ Take $\mathbf{k} = \hat{\mathbf{x}} + \mathbf{y}_j$, then
  $$\langle \mathbf{x} - \hat{\mathbf{x}}, \mathbf{y}_j \rangle \leq 0.$$
  - ⋄ Take $\mathbf{k} = \hat{\mathbf{x}} - \alpha_j \mathbf{y}_j$, then
  $$\langle \mathbf{x} - \hat{\mathbf{x}}, -\alpha_j \mathbf{y}_j \rangle \leq 0.$$
  - ⋄ Together,
  $$\langle \mathbf{x} - \hat{\mathbf{x}}, \mathbf{y}_i \rangle \leq 0, \text{ with equality if } \alpha_i > 0.$$

- Rewrite in matrix form,
  - $$\begin{pmatrix} \langle \mathbf{y}_1, \mathbf{y}_1 \rangle & \langle \mathbf{y}_2, \mathbf{y}_1 \rangle & \dots & \langle \mathbf{y}_n, \mathbf{y}_1 \rangle \\ \langle \mathbf{y}_1, \mathbf{y}_2 \rangle & & & \\ \vdots & & \ddots & \vdots \\ \langle \mathbf{y}_1, \mathbf{y}_n \rangle & & & \langle \mathbf{y}_n, \mathbf{y}_n \rangle \end{pmatrix} \begin{pmatrix} \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_n \end{pmatrix} - \begin{pmatrix} \langle \mathbf{x}, \mathbf{y}_1 \rangle \\ \langle \mathbf{x}, \mathbf{y}_2 \rangle \\ \vdots \\ \langle \mathbf{x}, \mathbf{y}_n \rangle \end{pmatrix} = \begin{pmatrix} z_1 \\ z_2 \\ \vdots \\ z_n \end{pmatrix}.$$
  - $z_i \geq 0$ for all i.
  - $\alpha_i z_i = 0$ 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 $\mathbf{b} \in \mathbb{R}^m$, want to minimize

$$\mathbf{x}^T Q \mathbf{x},$$

  subject to $A\mathbf{x} = \mathbf{b}$.

- The feasible set is a linear variety,

$$A\mathbf{x} = \mathbf{b} \Longleftrightarrow \mathbf{x} \in \mathbf{x}_0 + \mathcal{N}(A).$$

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

- This is a minimum norm problem.

  ◇ $\langle \mathbf{x}, \mathbf{y} \rangle = \mathbf{x}^T Q \mathbf{y}$ 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.