Chapter 1

Introduction

Contents

EECS 600
Function Space Methods in System Theory
Lecture Notes
J. Fessler

1.1

Motivation

Preface. "The primary objective of this book is to demonstrate that a rather large segment of the field of optimization can be effectively unified by a few geometric principles..."

p.2. "Most of the principal results in functional analysis are expressed as abstractions of intuitive geometric properties of ordinary 3D space."

Examples: point, line, plane, sphere, length, distance, angle, orthogonality, perpendicular projections, ...

"Some readers may look with great expectation towards functional analysis, hoping to discover new powerful techniques that will enable them to solve important problems beyond the reach of simpler mathematical analysis. Such hopes are rarely realized in practice. The primary utility of functional analysis ... is its role as a unifying discipline..."

The *unification* of ideas you have probably previously seen as special cases (such as the Cauchy-Schwarz inequality) is perhaps the main benefit of this course.

Keywords

- vector space, norm, inner product, Banach space, Hilbert space
- orthogonality, convexity, duality, hyperplanes, adjoints, pseudoinverse
- Euler-Lagrange equations, Gateaux and Fréchet differentials
- Kuhn-Tucker conditions, Lagrange multipliers, projections, etc.

1.2

Applications (of optimization)

- resource allocation
- planning (*e.g.*, production)
  Example. burn rate of booster rocket
- control
- approximation
- estimation (important for statistical signal processing)
- game theory

Wavelets

The concepts in this course will be of particular interest to students using tools like wavelets. For example, the second chapter of Vetterli and Kovacevic's text "Wavelets and subband coding" has a review of Hilbert spaces. Likewise, the classic book by Daubechies on wavelets contains many uses of function space methods. (Wavelets are not a prerequisite for EECS 600. But the topics of EECS 600 are nearly a prerequisite for serious analyses using wavelets!)

1.3

Main principles

0. Infinite-dimensional vector spaces
   Example. The space of continuous real-valued functions on an interval $[a, b]$:
$$C[a, b] = \{f : [a, b] \to \mathbb{R} : f \text{ is continuous}\}.$$

1. The **projection theorem**

   In 3D Euclidean space, the shortest distance from a point to a plane is determined by the **perpendicular** from the point to the plane

   (Picture) 2D, 3D cases showing point x, projection $x^\star$, and plane P

   Generalized to higher dimensions, this principle forms the basis for least-squares approximation, control, and estimation procedures!

   Example. Suppose we are given $f_0 \in \mathcal{L}_2[0, 1] \triangleq \left\{f : [0, 1] \to \mathbb{R} : \int_0^1 f^2(t)\,\mathrm{d}t < \infty\right\}$.
   Now consider the space of 2nd-order polynomials: $P_2 = \left\{at^2 + bt + c : a, b, c \in \mathbb{R}\right\}$.
   Problem. Find $p \in P_2$ that minimizes $\int_0^1 [f(t) - p(t)]^2\,\mathrm{d}t$.
   Solution. Interpret P_2 as a "plane" in an appropriate function space, and $\int_0^1 [\cdot - \cdot]^2\,\mathrm{d}t$ as a "distance." Then the projection theorem will say that the best p is such that $f - p \perp P$, which leads to a tractable, **finite-dimensional** linear algebra problem!

1b. **Completeness**

   If we use more and more components for approximation, will the error converge to zero? This depends in part on whether we have a **complete** basis. We claim in undergraduate signals and systems courses that "the Fourier basis is complete." Similar claims are made about various wavelet bases. This course provides the tools to make such statements rigorous.

2. The **Hahn-Banach theorem**

   In its simplest form, it states that given sphere and a point outside that sphere, there exists a (hyper)plane that separates them.

   (Picture) in 2D: · |○ ***Ch. 5 - over 100 pages away!***

3. **Duality**

   Converting between minimization and maximization problems, usually by exchanging vectors and hyperplanes.

   Example. The shortest distance from a point to a sphere equals the maximum of the distances from the point to separating hyperplanes. (Note that H.B. ensures the existence of such hyperplanes.)

   (Picture) in 2D: · /○

4. **Differentials**

   In ordinary calculus one maximizes a function by equating its derivative to zero (and then checking the sign of the second derivative).

   In finite dimensional spaces one equates the **gradient** of a function to zero.

   For optimization over infinite-dimensional spaces, we require generalizations of these concepts.

   Preview: at an extremum, the tangent hyperplane to the graph of the function is horizontal. So hyperplanes arise here too!

History

The mathematics of functional analysis is relatively new. Stefan **Banach**, the Polish founder of much of functional analysis, lived from 1892-1945 and published his book "Théorie des opérations linéaires" (which does not look like Polish to me) in 1932.

Only in 1973 did Enflo prove that there is a separable Banach space that has no Schauder basis (after Luenberger's book was written!) [1].

Prerequisites

- Linear algebra (vectors, matrices, eigenvalues, etc.) (Math 419 or equivalent)
  For example: the determinant of a (square) matrix is the product of its eigenvalues.
- Differentiation
  We say $f(t)$ is **differentiable** at t with **derivative** $\dot{f}(t)$ iff

$$\lim_{\delta \to 0} \frac{f(t+\delta) - f(t)}{\delta} = \dot{f}(t), \text{ i.e., } \forall \varepsilon > 0, \ \exists \delta_0 > 0 \text{ s.t. } |\delta| < \delta_0 \Longrightarrow \left| \frac{f(t+\delta) - f(t)}{\delta} - \dot{f}(t) \right| < \varepsilon.$$

- Limits and convergence
  A sequence x_n of real (or complex) numbers is said to **converge** to a limit x iff

$$\forall \varepsilon > 0, \ \exists N_\varepsilon < \infty \text{ s.t. } n \geq N_\varepsilon \Longrightarrow |x_n - x| < \varepsilon.$$

  Such concepts are discussed at length in Math 451.
  This course will be mathematically "self contained" in that any such concepts needed from Math 451 will be briefly reviewed.
  But, review **supremum**, **infimum**, **limit superior**, **limit inferior**.
  Other examples: if $a_n \to a$ and $b_n \to b$, then $a_n + b_n \to a + b$.
- Completeness of Euclidean space.
  Any Cauchy sequence of real numbers converges. (Proven in Math 451; we will take this as given.)

Sequences

Sequences are denoted $x_1, x_2, \ldots$ or $\{x_n\}_{n=1}^{\infty}$ or $\{x_n\}$.

Definition. A real number s is called the **limit superior** of a sequence $\{x_n\}$ and denoted $s = \limsup_{n\to\infty} x_n$ iff

$$\forall \varepsilon > 0, \ \exists N \text{ s.t. } \forall n > N, \ x_n < s + \varepsilon$$

and

$$\forall \varepsilon > 0 \text{ and } M > 0, \ \exists n > M, \text{ s.t. } x_n > s - \varepsilon.$$

If $\{x_n\}$ is not bounded above, we write $\limsup x_n = \infty$.

Definition. The **limit inferior** is

$$\liminf_{n\to\infty} x_n = -\limsup_{n\to\infty} -x_n.$$

If $\liminf x_n = \limsup x_n = y$, then we write $\lim_{n\to\infty} x_n = y$.

<u>Example</u>. If $x_n = (-1)^n + 7/n$, then $\limsup x_n = 1$ and $\liminf x_n = -1$.

Sets of real numbers

Definition. The **least upper bound** or **supremum** of a set $\mathcal{S}$ of real numbers that is bounded above is the smallest real number y such that $x \leq y, \ \forall x \in \mathcal{S}$, and is denoted $\sup \{x : x \in \mathcal{S}\}$. If $\mathcal{S}$ is not bounded above, then $\sup \{x : x \in \mathcal{S}\} = \infty$.

Definition. Similarly, the **greatest lower bound** or **infimum** of $\mathcal{S}$ is denoted $\inf \{x : x \in \mathcal{S}\}$, which can be $-\infty$.

Fact. These quantities always exist (they may be $\pm\infty$) for a nonempty set $\mathcal{S}$.

<u>Example</u>. If $\mathcal{S} = \mathbb{Q} \cap (0, 1)$, then $\sup \mathcal{S} = 1$ and $\inf \mathcal{S} = 0$.

Mathematical notation

Symbol	Meaning
$\mathbb{R}$	Real numbers
$\mathbb{C}$	Complex numbers
$\mathbb{N}$	Natural numbers $\{1, 2, 3, \ldots\}$
$\mathbb{Z}$	Integers $\{\ldots, -2, -1, 0, 1, 2, \ldots\}$ (German Zahlen = integers)
$\mathbb{Q}$	rational numbers (quotients)

Symbol	how it is read
:	"such that"
$\exists$	"there exists"
$\forall$	"for all"
$\Longleftrightarrow$	"if and only if" (iff)
$\Longrightarrow$	"implies"
$\rightarrow$	"maps into"

Sets

This review of sets is provided for completeness, and contains more than we really need in 600.

Definitions

Set theory was largely created by G. Cantor (1845-1918), who, according to [2, p. 1] defined a set as follows.

Definition. A **set** is a collection of definite, distinguishable objects of our thought, to be conceived as a whole.

The "objects of our thought" in sets are called **elements** or **members**.

Often upper case letters early in the alphabet denote sets: A, B, C.
Often lower case letters denote generic elements of sets: a, b, x, y, s.
The **universal set**, denoted Ω, contains all the elements of interest in a particular context.
The **empty set**, denoted $\emptyset$, contains no elements.

Set membership

- $x \in A$ means x is an element of set A. Read: x is in A, or x belongs to A, or x is a member of A.
- $x \notin A$ means x is not an element of set A

Set notation

Often we enumerate set contents in braces: $A = \{1, 2, \ldots, 10\}$, or by properties: $A = \{x \in \mathbb{N} : x \leq 10\}$.
By definition: $A = \{x \in \Omega : x \in A\}$
An interval on the real line: $A = [2, 5) = \{x \in \mathbb{R} : 2 \leq x < 5\}$

Set relationships

Subset $A \subseteq B$ iff $\forall x,\ x \in A \Longrightarrow x \in B$ (*i.e.*, every element of A must also be in B)
Proper Subset $A \subset B$ iff $A \subseteq B$ and $\exists x \in B$ such that $x \notin A$
Equality $A = B$ iff ($\forall x,\ x \in A \iff x \in B$) iff ($A \subseteq B$ and $B \subseteq A$)
Disjoint or Mutually Exclusive Sets: no common elements: $x \in A \Longrightarrow x \notin B$. Equivalently: $x \in B \Longrightarrow x \notin A$

Set operations

Union	$A \cup B = A + B = \{x \in \Omega : x \in A \text{ or } x \in B\}$
Intersection	$A \cap B = AB = \{x \in \Omega : x \in A \text{ and } x \in B\}$
Complement	$A^c = \widetilde{A} = \{x \in \Omega : x \notin A\}$
Difference (aka **set reduction**)	$A - B = A \cap B^c = \{x \in A : x \notin B\}$ (*i.e.*, the elements of A that are *not* in B)

(Picture) Venn Diagram

Disjoint sets or Mutually exclusive sets

Sets A and B are **disjoint** iff $A \cap B = \emptyset$

Set algebra

Commutative Laws	$A \cup B = B \cup A$
	$A \cap B = B \cap A$
Associative Laws	$A \cup (B \cup C) = (A \cup B) \cup C$
	$A \cap (B \cap C) = (A \cap B) \cap C$
Distributive Laws	$A \cup (B \cap C) = (A \cup B) \cap (A \cup C)$
	$A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$
De Morgan's Laws	$\widetilde{A \cap B} = \widetilde{A} \cup \widetilde{B}$
	$\widetilde{A \cup B} = \widetilde{A} \cap \widetilde{B} \quad \star$

Proof of $\star$: $x \in \widetilde{A \cup B} \iff x \notin A \cup B \iff x \notin A$ and $x \notin B \iff x \in \widetilde{A}$ and $x \notin \widetilde{B} \iff x \in \widetilde{A} \cap \widetilde{B}$.

Because of the commutative and associative laws, the following notation is unambiguous:

$$\bigcup_{i=1}^{n} A_i = A_1 \cup A_2 \cup \ldots \cup A_n \quad \text{and} \quad \bigcap_{i=1}^{n} A_i = A_1 \cap A_2 \cap \ldots \cap A_n.$$

One can also have countable and uncountable unions. For example, by definition:

$$A = \bigcup_{x \in A} \{x\}.$$

Set partition

The sets $A_1, A_2, \ldots, A_n$ are said to **partition** B iff

- $B = \bigcup_{i=1}^{n} A_i$, and
- $A_i \cap A_j = \emptyset, \; i \neq j$ (*i.e.*, the A_i's are mutually exclusive or disjoint).

Duality principle

In any set identity, if you replace $\cap$ by $\cup$, $\cup$ by $\cap$, Ω by $\emptyset$, and $\emptyset$ by Ω, then the new identity is also true.
Example: Distributive Laws, De Morgan's Laws.
Example: $A \cap \Omega = A$ becomes $A \cup \emptyset = A$

Set properties (important for probability)

- $\widetilde{\emptyset} = \Omega, \; \widetilde{\Omega} = \emptyset$
- $A \cup \widetilde{A} = \Omega$
- $A \cap \widetilde{A} = \emptyset$.
- Thus A and $\widetilde{A}$ partition Ω
- $\widetilde{\widetilde{A}} = A$
- $\widetilde{A} = \Omega - A$
- $A \cup \Omega = \Omega$, so from the **duality principle**: **??**
- etc.

Set functions

The following definitions generalize the usual definitions of functions on the real line used in calculus.

$f : A \to B$ is a **function** from A to B iff $\forall x \in A,\ f(x) \in B$, *i.e.*, for every element x of A, the function f assigns a value denoted $f(x)$ that is a member of B.
We also say f is a **mapping** from A **into** B.
Calculus is concerned primarily with functions on the real line: $f : \mathbb{R} \to \mathbb{R}$, such as $f(x) = x^2$.
In probability we use more general functions.

Example. In a coin tossing experiment with 2 coin tosses, the sample space is $\Omega = \{HH, HT, TH, TT\}$, and we might be interested in the function $f : \Omega \to \mathbb{R}$ that returns the number of tails:

$$f(s) = \begin{cases} 0, & \text{if } s = \text{HH} \\ 1, & \text{if } s = \text{HT} \\ 1, & \text{if } s = \text{TH} \\ 2, & \text{if } s = \text{TT}. \end{cases}$$

When $f : A \to B$, the set A is called the **domain** of f.
We say f is **defined on** A.
The elements $f(x)$ are called the *values* of f.
The **range** of f is denoted $R(f) = \{f(x) : x \in A\}$ and is the set of values of f.
Note that $R(f) \subseteq B$.

If $E \subseteq A$, then $f(E) = \{f(x) : x \in E\}$ is called the **image** of E under the mapping f.
Thus $R(f) = f(A)$. Also $f(A) \subseteq B$.
However, if $f(A) = B$, then we say f maps A **onto** B. This is a stronger condition than "into."

If $E \subseteq B$, then $f^{-1}(E) = \{x \in A : f(x) \in E\}$ is called the **inverse image** of E under f.
In particular, if $y \in B$, then $f^{-1}(y) = f^{-1}(\{y\}) = \{x \in A : f(x) = y\}$.

If for every $y \in B$, $f^{-1}(y)$ is (at most) a single element of A, then f is called a **one-to-one** mapping of A into B.
Equivalently, f is a **one-to-one** mapping of A into B iff $x_1 \in A$ and $x_2 \in A$ and $x_1 \neq x_2 \Longrightarrow f(x_1) \neq f(x_2)$.
In other words, if x_1 and x_2 are two distinct elements of A, then f must map those two elements to different values in B.
If in addition, $f(A) = B$, then f is called a one-to-one mapping of A onto B.

Examples

$$\begin{aligned} A_1 &= \{1, 2, 3\} \\ A_2 &= \{-3, -2, -1, 1, 2, 3\} \\ B_1 &= \{2, 4, 8\} \\ B_2 &= \{2, 3, 4, 8\} \end{aligned}$$

The function $f_1(x) = 2^{|x|}$ maps A_2 onto B_1, and maps A_2 into B_2, because $f_1(A_2) = \{2, 4, 8\}$.

The function $f_2(x) = 2^x$ is a one-to-one mapping of A_1 onto B_1, and is a one-to-one mapping of A_1 into B_2.
Note that $f_2^{-1}(3) = \emptyset$, because there are no points x in A_1 (or A_2) for which $f_2(x) = 3$.

Set "size"

If there exists a one-to-one mapping of A onto B, then we say A and B have the same **cardinality** or the same **cardinal number**, or that A and B can be put into one-to-one correspondence. Sometimes written $A \sim B$.

Possible set "sizes" or cardinalities

- **Empty**: $A = \emptyset$
- **Finite**: $A \sim \{1, 2, \ldots, n\}$ for a positive integer n
- **Countably Infinite**: $A \sim \{1, 2, 3, \ldots\}$ (all positive integers)
- **Uncountable** or **Uncountably Infinite**: (if none of the above apply)

Roughly speaking, a set is countably infinite if there is a way to write down all of its elements as an (infinitely long) list.
(A list is like a mapping defined on the set of positive integers, where $f(1)$ is the first item on the list, $f(2)$ is the second item, etc.)

Examples

- The set $\{red, green, blue, yellow\}$ is finite.
- The set of even positive integers is countably infinite. (Let $f(n) = 2n$).
- The set of all integers is countably infinite.
  (Let $f(n) = (-1)^n$ ceil$[\frac{n-1}{2}]$, where ceil[x] denotes the smallest integer $\geq x$.)
- The set of positive fractions is countably infinite.
- The interval [0,1) on the real line is uncountably infinite!

The real numbers are uncountably infinite

Proof that $A = [0, 1)$ is an uncountable set

(Proof by contradiction, *i.e.*, assume that A is countable, and show the assumption leads to a logical contradiction, so we must then conclude that A is uncountable.)

Suppose A is countably infinite. Then every $x \in A$ is one of the elements of the collection $\{a^1, a^2, a^3, \ldots\}$, since by definition, if A is countable, then there exists a function f from the positive integers onto A.

Now write each a^n in its decimal expansion:

$$a^n = 0.u_1^n u_2^n u_3^n \cdots = \sum_{i=1}^{\infty} u_i^n 10^{-i},$$

where $u_i^n \in \{0, 1, \ldots, 9\}$.
Now we construct a new decimal number:

$$y = 0.v_1 v_2 v_3 \cdots,$$

where we define

$$v_i = \begin{cases} 1, & \text{if } u_i^i = 0, \\ 0, & \text{otherwise.} \end{cases}$$

Clearly $y \in [0, 1) = A$ by construction, but $y \neq a^n$ for all n, since y and a^n differ in the nth decimal place!
This contradicts the existence of a function f from the positive integers onto A. Thus A must be uncountable. □

Space

The concept of a **space** is used throughout this course (and throughout math). A space is more specific than a mere **set** since it involves both a set and certain properties that the members of the set are assumed or shown to satisfy. Many spaces are named after the individuals who investigated them, *e.g.*, Euclidean space, Banach space, and Hilbert space, a practice that gives little insight into the properties!

Bibliography

[1] P. Enflo. A counterexample to the approximation problem in Banach spaces. *Acta Math*, 130:309–17, 1973.

[2] I. J. Maddox. *Elements of functional analysis*. Cambridge, 2 edition, 1988.