Mathematical Preliminaries

0.1 Preliminary Notation

It is assumed that the reader is familiar with the notion of a set and its elementary operations, and with some basic logic operators, e.g.

$x \in A$: x is an element of the set A
$x \notin A$: x does not belong to A
$B \subset A$: B is a subset of A
$B \cap A$: intersection of B and A
$B \cup A$: union of B and A
$a \Rightarrow b$: a is true implies that b is true (b not true implies a is not true)
$a \Leftrightarrow b$: a is true iff (if and only if) b is true
$\forall$: symbol "for all"

Throughout the course, R will denote the set of all real numbers; C the set of all complex numbers and $R_+ = \{x \in R \mid x \geq 0\}$ (i.e., the set of all nonnegative real numbers). Similarly, Z will denote the set of all integers and Z_+ the set of all nonnegative integers.

0.1.1 Functions

Given two sets X and Y, we denote a function f by

$$f : X \to Y$$

to mean that, for every $x \in X$, f assigns *one and only one* element $f(x) \in Y$. X is called the *domain* of f and we say that f maps X to Y. We define

$$f(X) \triangleq \{f(x) \mid x \in X\}$$

as the *range* of f.

At times it is convenient to define a function explicitly, for example

$$t \mapsto \cos(t)$$

means "the function that maps t to $\cos(t)$".

- $f : X \to Y$ is *onto* if $F(X) = Y$.
- $f : X \to Y$ is *one-to-one* if $f(x_1) = f(x_2) \Rightarrow x_1 = x_2$
  (or equivalently, $x_1 \neq x_2 \Rightarrow f(x_1) \neq f(x_2)$).

If f is one-to-one, then it has an inverse that maps $f(X) \to X$, which is normally represented as f^{-1}.

0.2 Vector Spaces

0.2.1 Algebraic Aspects

Definition A ***field*** F is a set of elements called *scalars* together with two binary operations, *addition* (+) and *multiplication* (·) such that for all $\alpha, \beta, \gamma \in F$ the following properties hold:

(a) Closure. $\alpha \cdot \beta \in F, \alpha + \beta \in F$

(b) Commutativity. $\alpha \cdot \beta = \beta \cdot \alpha$, $\alpha + \beta = \beta + \alpha$

(c) Associativity. $\alpha + (\beta + \gamma) = (\alpha + \beta) + \gamma$, $\alpha \cdot (\beta \cdot \gamma) = (\alpha \cdot \beta) \cdot \gamma$

(d) Distribution. $\alpha \cdot (\beta + \gamma) = (\alpha \cdot \beta) + (\alpha \cdot \gamma)$

(e) Identity. There exists an *additive identity* $0 \in F$ and a *multiplicative identity* $1 \in F$ such that $\alpha + 0 = \alpha$, $\alpha \cdot 1 = \alpha$

(f) Inverses. For all $\alpha \in F$ there exists an additive inverse $-\alpha \in F$ such that $\alpha + (-\alpha) = 0$. For all $\alpha \in F$, $\alpha \neq 0$ and a *multiplicative inverse* $\alpha^{-1} \in F$ such that $\alpha \cdot \alpha^{-1} = 1$

Examples The following are examples of fields:

- R = the set of real numbers
- C = the set of complex numbers
- Q = the set of rational numbers
- $R(s)$ = the set of rational functions in s with real coefficients

These are *not* fields:

- $R[s]$ = the set of polynomials in s with real coefficients. Why?
- R^{2x2} the set of real 2×2 matrices. Why?

Definition A ***vector space*** (V, F) is a set of *vectors* V together with a field F and two operations *vector-vector addition* (+) and *vector-scalar multiplication* (∘) such that for all $\alpha, \beta \in F$ and all $v_1, v_2, v_3 \in V$, the following properties hold:

(a) Closure. $v_1 + v_2 \in V$, $\alpha \circ v_1 \in V$

(b) Commutativity. $v_1 + v_2 = v_2 + v_1$

(c) Associativity. $(v_1 + v_2) + v_3 = v_1 + (v_2 + v_3)$

(d) Distribution. $\alpha \circ (\beta \circ v_1) = (\alpha \cdot \beta) \circ v_1$, $\alpha \circ (v_1 + v_2) = \alpha \circ v_1 + \alpha \circ v_2$

(e) Additive Identity. There exists a vector $0 \in V$ such that $v + 0 = v$ for all $v \in V$

(f) Additive Inverse. For all $v \in V$, there exists a $(-v) \in V$ such that $v + (-v) = 0$

We shall henceforth suppress the cumbersome notation · , ∘ as the appropriate action will be clear from context. Also, we shall often refer to a vector space V without explicit reference to the base field F (which will exclusively be R or C). We should however caution the reader that different choices of the base field F result in fundamentally different vector spaces (see example below).

Examples The following are examples of vector spaces:

- (R, R), (C, C) with addition and multiplication as defined in the field. Any field is a vector space over itself.
- (R^n, R), (C^n, C) with component-wise addition and scalar multiplication.
- ($R[s]$, R) with formal addition and scalar multiplication of polynomials.
- (C, R) is a vector space. Note that this vector space is fundamentally different from (C, C). Why?
- The space of infinite sequences of real numbers $x = (x_1, x_2, \ldots)$ with $x_i \in R$ on the field R.
- $C[a, b] = \{f : [a, b] \rightarrow R,\ f$ is continuous $\}$ (i.e., the set of all continuous functions which map the interval $[a, b] \subset R$ to R) is a vector space over R with pointwise addition and multiplication.
- The Lebesgue spaces $L_p\,[a,b]$, $1 \le p < \infty$ defined as

$$L_p[a,b] = \left\{ f : [a,b] \rightarrow R, \quad \int_a^b |f(t)|^p\,dt < \infty \right\}$$

are vector spaces over R with pointwise addition and multiplication.
(We will later talk more about L_p vector spaces).

(R, C) is *not* a vector space with the usual complex arithmetic. Why?

Definitions A set (possibly infinite) $S = \{v_i : i \in I\}$ of vectors from V is called ***linearly dependent*** if there exist scalars α_i, *not all zero* and only *finitely* many α_i being *nonzero* such that

$$\sum_{i \in I} \alpha_i v_i = 0$$

otherwise, the set of vectors S is said to be ***linearly independent***.

The *dimension* of a vector space V is the maximal number of linearly independent vectors in V.

A set B of vectors in V is called a *basis* for V if every vector in V can be *uniquely* expressed as a finite linear combination of vectors in B.

Basis are *not* unique.

Examples

-- In the vector space (R^2, R),

$$v_1 = \begin{bmatrix} 1 \\ 2 \end{bmatrix}, \ v_2 = \begin{bmatrix} 2 \\ 3 \end{bmatrix}, \ v_3 = \begin{bmatrix} 3 \\ 4 \end{bmatrix}$$

is a set of linearly dependent vectors because $-v_1 + 2v_2 + v_3 = 0$.

--

$$v_1 = \begin{bmatrix} \dfrac{1}{s+1} \\ \dfrac{1}{s+2} \end{bmatrix}, \ v_2 = \begin{bmatrix} \dfrac{s+2}{s^2+4s+3} \\ \dfrac{1}{s+3} \end{bmatrix}$$

are linearly dependent in (R^2 (s), R (s)), but linearly independent in (R^2 (s), R). Why?

-- The set of vectors $S = (1, t, t^2, \ldots)$ are linearly independent in C [0,1].

-- The dimension of (R^n, R) is n and the set

$$B = \left\{ \begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ \vdots \\ 0 \end{bmatrix}, \cdots, \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 1 \end{bmatrix} \right\} = \{ e_1, \quad e_2, \quad \cdots \quad e_n \}$$

qualifies as a basis for this vector space.

-- The dimension of (C^n, R) is $2n$. Exhibit a basis for this space.

Theorem *Let V be an n-dimensional vector space and let B be a collection of vectors drawn from V. then, B is a basis if and only if B contains n linearly independent vectors.*

Definition Let V be a vector space. A subset $W \subseteq V$ is called a *subspace* if W is itself a vector space.

Let $S = \{v_i : i \in I\}$ be a set of vectors drawn from V. The *span* of S is the set of all finite linear combinations of vectors in S. We will denote this set $SP(S)$.

Notice that S is a basis of V if

- S is a linearly independent set, and
- $SP(S) = V$.

Theorem A set $W \subseteq V$ *is a subspace if and only if it is closed under vector addition and scalar multiplication, i.e.*

$$\alpha w_1 \in W, \qquad w_1 + w_2 \in W$$

for all $\alpha \in F, \; w_1, w_2 \in W$.

It is clear that the zero vector must lie in every (nonempty) subspace.

Using the above result, it is easy to verify that given a set of vectors S, $SP\{S\}$ is a subspace.

0.2.2 Normed Vector Spaces

In the sequel, we consider vector spaces over the field C of complex numbers or the field R of real numbers.

Definition Let V be a vector space. A ***norm*** on V is a function $\| \cdot \| : V \to R$ such that

(a) $\| v \| \geq 0$ and $\| v \| = 0 \Leftrightarrow v = 0$.

(b) $\| \alpha v \| = | \alpha | \, \| v \|$ for all $\alpha \in F, \; v \in V$.

(c) $\| v_1 + v_2 \| \leq \| v_1 \| + \| v_2 \|$. (Triangle inequality)

A vector space on which a norm has been defined is called a ***normed space***.

Examples The following are examples of norms.

(i) On R^n or C^n :

-- $\| v \|_1 = \sum_{i=1}^n | v_i |$

-- $\| v \|_2 = \left(\sum_{i=1}^n | v_i |^2 \right)^{1/2}$ (Euclidean norm)

-- $\| v \|_p = \left(\sum_{i=1}^n | v_i |^p \right)^{1/p}$, $1 \le p < \infty$ (p-norm)

-- $\| v \|_\infty = \max_i | v_i |$ (sup norm)

where

$$v = \begin{bmatrix} v_1 \\ v_2 \\ \cdot \\ \cdot \\ v_n \end{bmatrix}$$

(ii) On the space of infinite sequences of real numbers $x = (x_1, x_2, \cdots)$ with $x_i \in R$ on the field R. Frequently used norms on the subspaces l_1 , l_p and l_∞ are respectively:

-- $\| x \|_1 = \sum_{i=1}^{\infty} | x_i |$

-- $\| x \|_p = \left(\sum_{i=1}^{\infty} |x_i|^p \right)^{1/p}$, $1 \le p < \infty$

-- $\| x \|_\infty = \sup_{i \ge i} | x_i |$

(iii) On the function spaces $L_1 [a,b]$, $L_p [a,b]$, $L_\infty [a,b]$:

-- $\| f \|_1 = \int_a^b |f(t)| dt$

-- $\| f \|_p = \left(\int_a^b | f(t) |^p dt \right)^{1/p}$ $1 \le p < \infty$

-- $\| f \|_\infty = \sup_{t \in [a,b]} | f(t) |$,

By sup we mean the essential supremum, i.e.,

$$\sup_{t \in [a,b]} | f(t) | = \inf \{ a: \quad | f(t) | < a \quad \text{almost everywhere} \}$$

(except on a set of measure zero).

(iv) On the space of matrices $R^{n\times n}$ or $C^{n\times n}$:

-- The *induced 2-norm*

$$\| M \|_2 = \sup_{x\neq 0} \frac{\| Mx \|_2}{\| x \|_2} = \sqrt{\lambda_{\max}(M^T M)} = \sigma_{\max}(M)$$

-- Other *induced norms*

$$\| M \|_\infty = \sup_{x\neq 0} \frac{\|Mx\|_\infty}{\|x\|_\infty} = \max_i \sum_j |m_{ij}| \quad \text{(row sum)}$$

$$\| M \|_1 = \sup_{x\neq 0} \frac{\|Mx\|_1}{\|x\|_1} = \max_j \sum_i |m_{ij}| \quad \text{(column sum)}$$

-- The *Frobenius norm*

$$\| M \|_F = \left[\sum_{1\le i,j\le n} | m_{i,j} |^2 \right]^{1/2} = \left[trace\,(M^T M) \right]^{1/2}$$

where $m_{i,j}$ is the i, j-th element of the matrix M.

0.2.3 Equivalent Norms

Definition Let V be a vector space on C. Let $\| \cdot \|_a$ and $\| \cdot \|_b$ be two norms on V. The norms $\| \cdot \|_a$ and $\| \cdot \|_b$ are said to be **equivalent** if there exist two positive numbers m_l and m_u such that

$$m_l \| v \|_a \le \| v \|_b \le m_u \| v \|_a , \quad \text{for all } v \in V.$$

Theorem *All norms on finite dimensional spaces (e.g., R^n, C^n) are equivalent.*

Example On R^n:

$$\| v \|_\infty \le \| v \|_2 \le \sqrt{n} \| v \|_\infty$$
$$\| v \|_\infty \le \| v \|_1 \le \sqrt{n} \| v \|_\infty$$
$$\frac{1}{n} \| v \|_1 \le \| v \|_2 \le \| v \|_1$$

0.2.4 Relations Between Normed Spaces

Norms in infinite dimensional spaces are generally not equivalent. However, several important relations can be established.

Theorem *On the normed spaces l_1, l_p and l_∞, which are subspaces of the set of infinite sequences of real numbers $x = (x_1, x_2, \cdots)$ with $x_i \in R$ on the field R,*

$$l_1 \subset l_p \subset l_\infty , \qquad 1 < p < \infty .$$

Proof:

- Consider any integer $p \in [1, \infty]$. If $x \in l_p$, then $\sum_{i=1}^{\infty} |x_i|^p < \infty$. Hence, $x \in l_\infty$ and $l_p \subset l_\infty$ for $p \in [1, \infty]$.
- For any integer $N \geq 0$ and any integer $p \leq 1$,

$$\sum_{i=1}^{N} |x_i|^p \leq \left(\sum_{i=1}^{N} |x_i| \right)^p \leq \left(\|x\|_1 \right)^p.$$

Hence, as $N \to \infty$, if $x \in l_1$, $\|x\|_p \leq \|x\|_1 < \infty$. Thus, $l_1 \subset l_p$. ◊

Fact: If $f: R_+ \to R$ and $f \in L_1 \cap L_\infty$ (i.e., f belongs to both L_1 and L_∞), then $f \in L_p$ for $p \in [1, \infty]$.

Proof:

- Define the set $I = \{t \mid |f(t)| \geq 1\}$. Since $f \in L_1$, the Lebesgue measure of the set I is finite. This fact, together with the fact that $f \in L_\infty$ implies that $\int_I |f(t)|^p \, dt < \infty$.
- Defining now the complement of I, $I^c \triangleq \{t \mid |f(t)| < 1\}$,

$$\int_{I^c} |f(t)|^p dt \leq \int_{I^c} |f(t)| dt < \infty, \quad \text{for all } p \in [1, \infty].$$

The conclusion follows from these two observations. ◊

0.2.5 Inner Product Spaces

Definition Let V be a vector space on C. An *inner product* on V is a function $< \cdot , \cdot >: V \times V \to C$ such that

(a) $< v, w > = \overline{< w, v >}$

(b) $< v, \alpha w > = \alpha < v, w >$.

(c) $< v, w_1 + w_2 > = < v, w_1 > + < v, w_2 >$.

(d) $< v, v > \geq 0$, $< v, v > = 0 \Leftrightarrow v = 0$.

where $\overline{\alpha}$ denotes the complex conjugate of $\alpha \in C$.

A vector space on which an inner product has been defined is called an *inner product space*.

Examples The following are examples of inner products.

-- In R^n, $< v, w > = v^T w$

-- In C^n, $< v, w > = v^* w$
where v^* denotes the complex conjugate transpose of $v \in C^n$, i.e., $v^* = \overline{v}^T$,

-- In $L_2 [a, b]$,

$$< f, g > = \int_a^b f(t) g(t) \, dt$$

Theorem *Let V be an inner product space. Then*

$$\| v \| = < v, v >^{1/2}$$

qualifies as a norm on V.

The norm defined above is said to be *induced* by the inner product. In an inner product space, this is the natural norm to use.

Cauchy Schwartz Inequality

Let V be an inner product space. Then

$$|< v, w >| \leq \|v\|^{1/2} \, \|w\|^{1/2}$$

An immediate consequence of the Cauchy-Schwartz inequality applied to the inner product space $L_2[a, b]$ is

$$\int_a^b f(t) g(t) \, dt \leq \left[\int_a^b f^2(t) \, dt \right]^{1/2} \left[\int_a^b g^2(t) \, dt \right]^{1/2}$$

Definition In an inner product space V, two vectors v, w are said to be ***orthogonal*** if $< v, w > = 0$. This is often written as $v \perp w$. Further, v is orthogonal to the set of vectors S if $v \perp w$ for all $w \in S$. This is often written as $v \perp S$. A set of vectors S is called *orthogonal* if

$$v \perp w \text{ for all } v \neq w, \; v, w \in S$$

and is called *orthonormal* if in addition $\| v \| = 1$ for all $v \in S$.

We now provide a simple geometric interpretation of the inner product.

Let V be an inner-product space and fix $v \in$ V. Let $b \in V$ be a unit vector and consider the subspace $S = Span\{b\}$. Define

$$\hat{s} = < v, b > s$$

We wish to find an optimal approximation for v from the subspace S. More precisely we wish to solve the following problem:

$$\min_{s \in S} \| v - s \|$$

We have the following result:

Lemma The vector $\hat{s}$ above is the optimal approximation in S of v.

Thus $< v, b >$ may be regarded as the length of the *projection* of v on S.

0.3 Hermitian and Positive Definite Matrices

Definition A matrix $U \in C^{n\times n}$ is called *unitary* if $U^*U = I = UU^*$.

A real unitary matrix is called an *orthogonal* matrix.

Lemma Let $U\in C^{n\times n}$ be unitary and consider the Hilbert space C^n equipped with the usual inner product. Then,

(a) The columns of U form an orthonormal basis of C^n.

(b) $\| Ux \| = \| x \|$

(c) $< Ux, Uy > = < x, y >$

(d) $U^{-1} = U^*$.

Rotation matrices are unitary.

Definition A matrix $H \in C^{n\times n}$ is called *Hermitian* if $H = H^*$. Symmetric matrices are in particular Hermitian.

We will now prove several results regarding Hermitian matrices. These results also hold almost *verbatim* for symmetric matrices.

Theorem *The eigenvalues of a Hermitian matrix H are all real.* Δ

Theorem *A Hermitian matrix H has a full set of eigenvectors. Moreover, these eigenvectors form an orthogonal set. As a consequence, Hermitian matrices can be diagonalized by unitary transformations, i.e., there exists a unitary matrix U such that*

$$H = UDU^*$$

where D is a diagonal matrix whose entries are the (real) eigenvalues of H. Δ

Theorem *Let $H \in C^{n\times n}$ be Hermitian. Then,*

(a) $\|H\|_2 = \sup_{v\neq 0} \frac{\|Hv\|_2}{\|v\|_2} = \lambda_{\max}(H)$

(b) $\inf_{v\neq 0} \frac{v^* H v}{v^* v} = \lambda_{\min}(H)$ or $\lambda_{\min}(H) v^* v \le v^* H v \le \lambda_{\max}(H) v^* v, \quad \forall v$ Δ

Definition A matrix $P \in C^{n\times n}$ is called *positive-definite* and written as $P > 0$ if P is Hermitian and further,

$$v^* P v > 0, \quad \text{for all } 0 \neq v \in C^n$$

A matrix $P \in C^{n\times n}$ is called *positive-semi-definite* and written as P if P is Hermitian and further,

$$v^* P v \ge 0, \quad \text{for all } v \in C^n$$

Analogous are the notions of *negative* and *negative-semi*-definite matrices. ♦

Theorem *Let* $P \in C^{n\times n}$ *be Hermitian. The following are equivalent.*

(a) $P > 0$

(b) *All the eigenvalues of P are positive*.

(c) *All the leading principal minors of P are positive*. Δ

Example For the matrix

$$P = \begin{bmatrix} p_{11} & p_{12} & p_{13} \\ p_{21} & p_{22} & p_{23} \\ p_{31} & p_{32} & p_{33} \end{bmatrix}$$

the leading principal minors are

$$p_{11}, \qquad \text{Det}\left\{ \begin{bmatrix} p_{11} & p_{12} \\ p_{21} & p_{22} \end{bmatrix} \right\} \qquad \text{and} \qquad \text{Det}\,\{P\}.$$

Theorem Let $P \in C^{n\times n}$ be Hermitian. The following are equivalent.

(a) $P \geq 0$

(b) *All the eigenvalues of P are* ≥ 0.

(d) $P = N^*N$, *where N is any matrix*.

A principal minor test for positive-semi-definiteness is significantly more complicated. Δ

Lemma *Let* $0 < P \in C^{n\times n}$ *and let* $X \in C^{n\times m}$.

(a) $\| x \|^2 = x^* P x$ qualifies as a norm on C^n

(b) $X^* P X \geq 0$

(c) $X^* P X > 0$ *if and only if rank* $(X) = m$ Δ

Definition Let $0 \leq P \in C^{n\times n}$. We can then write $P = UDU^*$ where U is unitary. Define the *square-root* of P (written as $P^{1/2}$) by

$$P^{1/2} = UD^{1/2}U^*$$

It is evident that $P^{1/2}$ as defined above is Hermitian, and moreover $P^{1/2} \geq 0$. Furthermore, if $P > 0$, then $P^{1/2} > 0$. ♦