APPENDIX A

Some Linear Algebra

A.1 Matrices

The collection of m, n matrices

$$\boldsymbol{A} = \begin{pmatrix} a_{1,1}, & \cdots & , a_{1,n} \\ \cdots & & \cdots \\ a_{m,1}, & \cdots & , a_{m,n} \end{pmatrix}$$

with real elements $a_{i,j}$ is denoted by $\mathbb{R}^{m,n}$. If $n = 1$ then $\boldsymbol{A}$ is called a column vector. Similarly, if $m = 1$ then $\boldsymbol{A}$ is a row vector. We let $\mathbb{R}^m$ denote the collection of all column or row vectors with m real components.

A.1.1 Nonsingular matrices, and inverses.

Definition A.1. *A collection of vectors $\boldsymbol{a}_1, \ldots, \boldsymbol{a}_n \in \mathbb{R}^m$ is* linearly independent *if $x_1\boldsymbol{a}_1 + \cdots + x_n\boldsymbol{a}_n = \boldsymbol{0}$ for some real numbers $x_1, \ldots, x_n$, implies that $x_1 = \cdots = x_n = 0$.*

Suppose $\boldsymbol{a}_1, \ldots, \boldsymbol{a}_n$ are the columns of a matrix $\boldsymbol{A} \in \mathbb{R}^{m,n}$. For a vector $\boldsymbol{x} = (x_1, \ldots, x_n)^T \in \mathbb{R}^n$ we have $\boldsymbol{A}\boldsymbol{x} = \sum_{j=1}^n x_j\boldsymbol{a}_j$. It follows that the collection $\boldsymbol{a}_1, \ldots, \boldsymbol{a}_n$ is linearly independent if and only if $\boldsymbol{A}\boldsymbol{x} = \boldsymbol{0}$ implies $\boldsymbol{x} = \boldsymbol{0}$.

Definition A.2. *A square matrix $\boldsymbol{A}$ such that $\boldsymbol{A}\boldsymbol{x} = \boldsymbol{0}$ implies $\boldsymbol{x} = \boldsymbol{0}$ is said to be nonsingular.*

Definition A.3. *A square matrix $\boldsymbol{A} \in \mathbb{R}^{n,n}$ is said to be invertible if for some $\boldsymbol{B} \in \mathbb{R}^{n,n}$*

$$\boldsymbol{B}\boldsymbol{A} = \boldsymbol{A}\boldsymbol{B} = \boldsymbol{I},$$

where $\boldsymbol{I} \in \mathbb{R}^{n,n}$ is the identity matrix.

An invertible matrix $\boldsymbol{A}$ has a unique inverse $\boldsymbol{B} = \boldsymbol{A}^{-1}$. If $\boldsymbol{A}, \boldsymbol{B}$, and $\boldsymbol{C}$ are square matrices, and $\boldsymbol{A} = \boldsymbol{B}\boldsymbol{C}$, then $\boldsymbol{A}$ is invertible if and only if both $\boldsymbol{B}$ and $\boldsymbol{C}$ are also invertible. Moreover, the inverse of $\boldsymbol{A}$ is the product of the inverses of $\boldsymbol{B}$ and $\boldsymbol{C}$ in reverse order, $\boldsymbol{A}^{-1} = \boldsymbol{C}^{-1}\boldsymbol{B}^{-1}$.

A.1.2 Determinants.

The determinant of a square matrix $\boldsymbol{A}$ will be denoted $\det(\boldsymbol{A})$ or

$$\begin{vmatrix} a_{1,1}, & \cdots & , a_{1,n} \\ \vdots & & \vdots \\ a_{n,1}, & \cdots & , a_{n,n} \end{vmatrix}.$$

Recall that the determinant of a 2×2 matrix is

$$\begin{vmatrix} a_{1,1} & a_{1,2} \\ a_{2,1} & a_{2,2} \end{vmatrix} = a_{1,1}a_{2,2} - a_{1,2}a_{2,1}.$$

A.1.3 Criteria for nonsingularity and singularity.

We state without proof the following criteria for nonsingularity.

Theorem A.4. *The following is equivalent for a square matrix* $\boldsymbol{A} \in \mathbb{R}^{n,n}$.

1. $\boldsymbol{A}$ *is nonsingular.*
2. $\boldsymbol{A}$ *is invertible.*
3. $\boldsymbol{A}\boldsymbol{x} = \boldsymbol{b}$ *has a unique solution* $\boldsymbol{x} = \boldsymbol{A}^{-1}b$ *for any* $\boldsymbol{b} \in \mathbb{R}^n$.
4. $\boldsymbol{A}$ *has linearly independent columns.*
5. $\boldsymbol{A}^T$ *is nonsingular.*
6. $\boldsymbol{A}$ *has linearly independent rows.*
7. $\det(\boldsymbol{A}) \neq \boldsymbol{0}$.

We also have a number of criteria for a matrix to be singular.

Theorem A.5. *The following is equivalent for a square matrix* $\boldsymbol{A} \in \mathbb{R}^{n,n}$.

1. *There is a nonzero* $\boldsymbol{x} \in \mathbb{R}^n$ *so that* $\boldsymbol{A}\boldsymbol{x} = \boldsymbol{0}$.
2. $\boldsymbol{A}$ *has no inverse.*
3. $\boldsymbol{A}\boldsymbol{x} = \boldsymbol{b}$ *has either no solution or an infinite number of solutions.*
4. $\boldsymbol{A}$ *has linearly dependent columns.*
5. *There is a nonzero* $\boldsymbol{x}$ *so that* $\boldsymbol{x}^T\boldsymbol{A} = \boldsymbol{0}$.
6. $\boldsymbol{A}$ *has linearly dependent rows.*
7. $\det(\boldsymbol{A}) = \boldsymbol{0}$.

Corollary A.6. *A matrix with more columns than rows has linearly dependent columns.*

Proof. Suppose $\boldsymbol{A} \in \mathbb{R}^{m,n}$ with $n > m$. By adding $n - m$ rows of zeros to $\boldsymbol{A}$ we obtain a square matrix $\boldsymbol{B} \in \mathbb{R}^{n,n}$. This matrix has linearly dependent rows. By Theorem A.4 the matrix $\boldsymbol{B}$ has linearly dependent columns. But then the columns of $\boldsymbol{A}$ are also linearly dependent. ∎

A.2 Vector Norms

Formally, a *vector norm* $|| \; || = ||\boldsymbol{x}||$, is a function $|| \; || : \mathbb{R}^n \to [0, \infty)$ that satisfies for $\boldsymbol{x}, \boldsymbol{y}, \in \mathbb{R}^n$, and $\alpha \in \mathbb{R}$ the following properties

$$\begin{array}{ll} 1. & ||\boldsymbol{x}|| = 0 \quad \text{implies} \quad \boldsymbol{x} = \boldsymbol{0}. \\ 2. & ||\alpha \boldsymbol{x}|| = |\alpha| ||\boldsymbol{x}||. \\ 3. & ||\boldsymbol{x} + \boldsymbol{y}|| \le ||\boldsymbol{x}|| + ||\boldsymbol{y}||. \end{array} \tag{A.1}$$

Property 3 is known as the *Triangle Inequality.* For us the most useful class of norms are the p or ℓ^p norms. They are defined for $p \ge 1$ and $\boldsymbol{x} = (x_1, x_2, \ldots, x_n)^T \in \mathbb{R}^n$ by

$$\begin{aligned} ||\boldsymbol{x}||_p &= (|x_1|^p + |x_2|^p + \cdots + |x_n|^p)^{1/p}. \\ ||\boldsymbol{x}||_\infty &= \max_i |x_i|. \end{aligned} \tag{A.2}$$

Since

$$||\boldsymbol{x}||_\infty \le ||\boldsymbol{x}||_p \le n^{1/p} ||\boldsymbol{x}||_\infty, \qquad p \ge 1 \tag{A.3}$$

and $\lim_{p\to\infty} n^{1/p} = 1$ for any $n \in \mathbb{N}$ we see that $\lim_{p\to\infty} ||\boldsymbol{x}||_p = ||\boldsymbol{x}||_\infty$.

The 1,2, and ∞ norms are the most important. We have

$$||\boldsymbol{x}||_2^2 = x_1^2 + \cdots + x_n^2 = \boldsymbol{x}^T \boldsymbol{x}. \tag{A.4}$$

Lemma A.7 (The Hölder inequality)**.** *We have for $1 \le p \le \infty$ and $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}$*

$$\sum_{i=1}^n |x_i y_i| \le ||\boldsymbol{x}||_p ||\boldsymbol{y}||_q, \quad \textit{where} \quad \frac{1}{p} + \frac{1}{q} = 1. \tag{A.5}$$

Proof. We base the proof on properties of the exponential function. Recall that the exponential function is convex, i.e. with $f(x) = e^x$ we have the inequality

$$f(\lambda x + (1 - \lambda) y) \le \lambda f(x) + (1 - \lambda) f(y) \tag{A.6}$$

for every $\lambda \in [0, 1]$ and $x, y \in \mathbb{R}$.

If $\boldsymbol{x} = \boldsymbol{0}$ or $\boldsymbol{y} = \boldsymbol{0}$, there is nothing to prove. Suppose $\boldsymbol{x}, \boldsymbol{y} \ne \boldsymbol{0}$. Define $\boldsymbol{u} = \boldsymbol{x}/||\boldsymbol{x}||_p$ and $\boldsymbol{v} = \boldsymbol{y}/||\boldsymbol{y}||_q$. Then $||\boldsymbol{u}||_p = ||\boldsymbol{v}||_q = 1$. If we can prove that $\sum_i |u_i v_i| \le 1$, we are done because then $\sum_i |x_i y_i| = ||\boldsymbol{x}||_p ||\boldsymbol{y}||_q \sum_i |u_i v_i| \le ||\boldsymbol{x}||_p ||\boldsymbol{y}||_q$. Since $|u_i v_i| = |u_i||v_i|$, we can assume that $u_i \ge 0$ and $v_i \ge 0$. Moreover, we can assume that $u_i > 0$ and $v_i > 0$ because a zero term contributes no more to the left hand side than to the right hand side of (A.5). Let s_i, t_i be such that $u_i = e^{s_i/p}$, $v_i = e^{t_i/q}$. Taking $f(x) = e^x, \lambda = 1/p$, $1 - \lambda = 1/q$, $x = s_i$ and $y = t_i$ in (A.6) we find

$$e^{s_i/p + t_i/q} \le \frac{1}{p} e^{s_i} + \frac{1}{q} e^{t_i}.$$

But then

$$\sum_i |u_i v_i| = \sum_i e^{s_i/p + t_i/q} \le \frac{1}{p} \sum_i e^{s_i} + \frac{1}{q} \sum_i e^{t_i} = \frac{1}{p} \sum_i u_i^p + \frac{1}{q} \sum_i v_i^q = \frac{1}{p} + \frac{1}{q} = 1.$$

This completes the proof of (A.5). ∎

When $p = 2$ then $q = 2$ and the Hölder inequality is associated with the names Buniakowski-Cauchy-Schwarz.

Lemma A.8 (The Minkowski inequality)**.** *We have for* $1 \le p \le \infty$ *and* $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}$

$$||\boldsymbol{x} + \boldsymbol{y}||_p \le ||\boldsymbol{x}||_p + ||\boldsymbol{y}||_p. \tag{A.7}$$

Proof. Let $\boldsymbol{u} = (u_1, \dots, u_n)$ with $u_i = |x_i + y_i|^{p-1}$. Since $q(p-1) = p$ and $p/q = p-1$, we find

$$||\boldsymbol{u}||_q = (\sum_i |x_i + y_i|^{q(p-1)})^{1/q} = (\sum_i |x_i + y_i|^p)^{1/q} = ||\boldsymbol{x} + \boldsymbol{y}||_p^{p/q} = ||\boldsymbol{x} + \boldsymbol{y}||_p^{p-1}.$$

Using this and the Hölder inequality we obtain

$$\begin{aligned} ||\boldsymbol{x} + \boldsymbol{y}||_p^p = \sum_i |x_i + y_i|^p &\le \sum_i |u_i||x_i| + \sum_i |u_i||y_i| \le (||\boldsymbol{x}||_p + ||\boldsymbol{y}||_p)||\boldsymbol{u}||_q \\ &\le (||\boldsymbol{x}||_p + ||\boldsymbol{y}||_p)||\boldsymbol{x} + \boldsymbol{y}||_p^{p-1}. \end{aligned}$$

Dividing by $||\boldsymbol{x} + \boldsymbol{y}||_p^{p-1}$ proves Minkowski. ∎

Using the Minkowski inequality it follows that the p norms satisfies the axioms for a vector norm.

In (A.3) we established the inequality

$$||\boldsymbol{x}||_\infty \le ||\boldsymbol{x}||_p \le n^{1/p}||\boldsymbol{x}||_\infty, \quad p \ge 1.$$

More generally, we say that two vector norms $||\ ||$ and $||\ ||'$ are *equivalent* if there exists positive constants μ and M such that

$$\mu||\boldsymbol{x}|| \le ||\boldsymbol{x}||' \le M||\boldsymbol{x}|| \tag{A.8}$$

for all $\boldsymbol{x} \in \mathbb{R}^n$.

Theorem A.9. *All vector norms on* $\mathbb{R}^n$ *are equivalent.*

Proof. It is enough to show that a vector norm $||\ ||$ is equivalent to the l_∞ norm, $||\ ||_\infty$. Let $\boldsymbol{x} \in \mathbb{R}^n$ and let $\boldsymbol{e}_i, i = 1, \dots, n$ be the unit vectors in $\mathbb{R}^n$. Writing $\boldsymbol{x} = x_1\boldsymbol{e}_1 + \dots + x_n\boldsymbol{e}_n$ we have

$$||\boldsymbol{x}|| \le \sum_i |x_i|||\boldsymbol{e}_i|| \le ||\boldsymbol{x}||_\infty M, \quad M = \sum_i ||\boldsymbol{e}_i||.$$

To find $\mu > 0$ such that $||\boldsymbol{x}|| \ge \mu||\boldsymbol{x}||_\infty$ for all $\boldsymbol{x} \in \mathbb{R}^n$ is less elementary. Consider the function f given by $f(x) = ||\boldsymbol{x}||$ defined on the l_∞ "unit ball"

$$S = \{\boldsymbol{x} \in \mathbb{R}^n : ||\boldsymbol{x}||_\infty = 1\}.$$

S is a closed and bounded set. From the inverse triangle inequality

$$|\,||\boldsymbol{x}|| - ||\boldsymbol{y}||\,| \le ||\boldsymbol{x} - \boldsymbol{y}||, \quad \boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^n.$$

it follows that f is continuous on S. But then f attains its maximum and minimum on S, i.e. there is a point $\boldsymbol{x}^* \in S$ such that

$$||\boldsymbol{x}^*|| = \min_{\boldsymbol{x} \in S} ||\boldsymbol{x}||.$$

Moreover, since $\boldsymbol{x}^*$ is nonzero we have $\mu := ||\boldsymbol{x}^*|| > 0$. If $\boldsymbol{x} \in \mathbb{R}^n$ is nonzero then $\boldsymbol{x} = \boldsymbol{x}/||\boldsymbol{x}||_\infty \in S$. Thus

$$\mu \leq ||\boldsymbol{x}|| = ||\frac{\boldsymbol{x}}{||\boldsymbol{x}||_\infty}|| = \frac{1}{||\boldsymbol{x}||_\infty}||\boldsymbol{x}||,$$

and this establishes the lower inequality. ∎

It can be shown that for the p norms we have for any q with $1 \leq q \leq p \leq \infty$

$$||\boldsymbol{x}||_p \leq ||\boldsymbol{x}||_q \leq n^{1/q-1/p}||\boldsymbol{x}||_p, \quad \boldsymbol{x} \in \mathbb{R}^n. \tag{A.9}$$

<

A.3 Vector spaces of functions

In $\mathbb{R}^m$ we have the operations $\boldsymbol{x} + \boldsymbol{y}$ and $a\boldsymbol{x}$ of vector addition and multiplication by a scalar $a \in \mathbb{R}$. Such operations can also be defined for functions. As an example, if $f(x) = x$, $g(x) = 1$, and a, b are real numbers then $af(x) + bg(x) = ax + b$. In general, if f and g are two functions defined on the same set I and $a \in \mathbb{R}$, then the sum $f + g$ and the product af are functions defined on I by

$$(f+g)(x) = f(x) + g(x),$$
$$(af(x) = af(x).$$

Two functions f and g defined on I are equal if $f(x) = g(x)$ for all $x \in I$. We say that f is the zero function, i.e. $f = 0$, if $f(x) = 0$ for all $x \in I$.

Definition A.10. *Suppose S is a collection of real valued or vector valued functions, all defined on the same set I. The collection S is called a vector space if $af + bg \in S$ for all $f, g \in S$ and all $a, b \in \mathbb{R}$. A subset T of S is called a subspace of S if T itself is a vector space.*

Example A.11. Vector spaces

- All polynomials π_d of degree at most d.
- All polynomials of all degrees.
- All trigonometric polynomials $a_0 + \sum_{k=1}^{d}(a_k \cos kx + b_k \sin kx$ of degree at most d.
- The set $C(I)$ of all continuous real valued functions defined on I.
- The set $C^r(I)$ of all real valued functions defined on I with continuous j'th derivative for $j = 0, 1, \ldots, r$.

Definition A.12. *A vector space S is said to be* finite dimesional *if*

$$S = \text{span}(\phi_1, \dots, \phi_n) = \{\sum_{j=1}^{n} c_j\phi_j : c_j \in \mathbb{R}\},$$

for a finite number of functions $\phi_1, \dots, \phi_n$ in S. The functions $\phi_1, \dots, \phi_n$ are said to span *or* generate S.

Of the examples above the space $\pi_d = \text{span}(1, x, x^2, \dots x^d)$ generated by the monomials $1, x, x^2, \dots x^d$ is finite dimensional. Also the trigonometric polynomials are finite dimensional. The space of all polynomials of all degrees is not finite dimensional. To see this we observe that any finite set cannot generate the monomial x^{d+1} where d is the maximal degree of the elements in the spanning set. Finally we observe that $C(I)$ and $C^r(I)$ contain the space of polynomials of all degrees as a subspace. Hence they are not finite dimensional,

If $f \in S = \text{span}(\phi_1, \dots, \phi_n)$ then $f = \sum_{j=1}^{n} c_j\phi_j$ for some $\boldsymbol{c} = (c_1, \dots, c_n)$. With $\boldsymbol{\phi} = (\phi_1, \dots, \phi_n)^T$ we will often use the vector notation

$$f(x) = \boldsymbol{\phi}(x)^T\boldsymbol{c} \tag{A.10}$$

for f.

A.3.1 Linear independence and bases

All vector spaces in this section will be finite dimensional.

Definition A.13. *A set of functions $\boldsymbol{\phi} = (\phi_1, \dots, \phi_n)^T$ in a vector space S is said to be* linearly independent *on a subset J of I if $\boldsymbol{\phi}(x)^T\boldsymbol{c} = c_1\phi_1(x) + \dots + c_n\phi_n(x) = 0$ for all $x \in J$ implies that $\boldsymbol{c} = \boldsymbol{0}$. If $J = I$ then we simply say that $\boldsymbol{\phi}$ is linearly independent.*

If $\boldsymbol{\phi}$ is linearly independent then the representation in (A.10) is unique. For if $f = \boldsymbol{\phi}^T\boldsymbol{c} = \boldsymbol{\phi}^T\boldsymbol{b}$ for some $\boldsymbol{c}, \boldsymbol{b} \in \mathbb{R}^n$ then $f = \boldsymbol{\phi}^T(\boldsymbol{c} - \boldsymbol{b}) = \boldsymbol{0}$. Since $\boldsymbol{\phi}$ is linearly independent we have $\boldsymbol{c} - \boldsymbol{b} = \boldsymbol{0}$, or $\boldsymbol{c} = \boldsymbol{b}$.

Definition A.14. *A set of functions $\boldsymbol{\phi}^T = (\phi_1, \dots, \phi_n)$ in a vector space S is* a basis for *S if the following two conditions hold*

1. *$\boldsymbol{\phi}$ is linearly independent.*
2. *$S = \text{span}(\boldsymbol{\phi})$.*

Theorem A.15. *The monomials $1, x, x^2, \dots x^d$ are linearly independent on any set $J \subset \mathbb{R}$ containing at least $d+1$ distinct points. In particular these functions form as basis for π_d.*

Proof. Let $x_0, \dots, x_d$ be $d+1$ distinct points in J, and let $p(x) = c_0 + c_1x + \dots + c_dx^d = 0$ for all $x \in J$. Then $p(x_i) = 0$, for $i = 0, 1, \dots, d$. Since a nonzero polynomial of degree d can have at most d zeros we conclude that p must be the zero polynomial. But then $c_k = p^{(k)}(0)/k! = 0$ for $k = 0, 1, \dots, d$. It follows that the monomial is a basis for π_d since they span π_d by definition. ∎

To prove some basic results about bases in a vector space of functions it is convenient to introduce a matrix transforming one basis into another.

Lemma A.16. *Suppose S and T are finite dimensional vector spaces with $S \subset T$, and let $\boldsymbol{\phi} = (\phi_1, \ldots, \phi_n)^T$ be a basis for S and $\boldsymbol{\psi} = (\psi_1, \ldots, \psi_m)^T$ a basis for T. Then*

$$\boldsymbol{\phi} = \boldsymbol{A}^T \boldsymbol{\psi}, \tag{A.11}$$

for some matrix $\boldsymbol{A} \in \mathbb{R}^{m,n}$. If $f = \boldsymbol{\phi}^T \boldsymbol{c} \in S$ is given then $f = \boldsymbol{\psi}^T \boldsymbol{b}$ with

$$\boldsymbol{b} = \boldsymbol{A}\boldsymbol{c}. \tag{A.12}$$

Moreover $\boldsymbol{A}$ has linearly independent columns.

Proof. Since $\phi_j \in T$ there are real numbers $a_{i,j}$ such that

$$\phi_j = \sum_{i=1}^{m} a_{i,j} \psi_i, \quad \text{for} \quad j = 1, \ldots, n,$$

This equation is simply the component version of (A.11). If $f \in S$ then $f \in T$ and $f = \boldsymbol{\psi}^T \boldsymbol{b}$ for some $\boldsymbol{b}$. By (A.11) we have $\boldsymbol{\phi}^T = \boldsymbol{\psi}^T \boldsymbol{A}$ and $f = \boldsymbol{\phi}^T \boldsymbol{c} = \boldsymbol{\psi}^T \boldsymbol{A}\boldsymbol{c}$ or $\boldsymbol{\psi}^T \boldsymbol{b} = \boldsymbol{\psi}^T \boldsymbol{A}\boldsymbol{c}$. Since $\boldsymbol{\psi}$ is linearly independent we get (A.12). Finally, to show that $\boldsymbol{A}$ has linearly independent columns suppose $\boldsymbol{A}\boldsymbol{c} = \boldsymbol{0}$. Define $f \in S$ by $f = \boldsymbol{\phi}^T \boldsymbol{c}$. By (A.11) we have $f = \boldsymbol{\psi}^T \boldsymbol{A}\boldsymbol{c} = \boldsymbol{0}$. But then $f = \boldsymbol{\phi}^T \boldsymbol{c} = \boldsymbol{0}$. Since $\boldsymbol{\phi}$ is linearly independent we conclude that $\boldsymbol{c} = \boldsymbol{0}$. ∎

The matrix $\boldsymbol{A}$ in Lemma A.16 is called a *change of basis matrix*.

A basis for a vector space generated by n functions can have at most n elements.

Lemma A.17. *If $\boldsymbol{\psi} = (\psi_1 \ldots, \psi_k)^T$ is a linearly independent set in a vector space $S =$ span$(\phi_1, \ldots, \phi_n)$, then $k \le n$.*

Proof. With $\boldsymbol{\phi} = (\phi_1, \ldots, \phi_n)^T$ we have

$$\boldsymbol{\psi} = \boldsymbol{A}^T \boldsymbol{\phi}, \quad \text{for some} \quad \boldsymbol{A} \in \mathbb{R}^{n,k}.$$

If $k > n$ then $\boldsymbol{A}$ is a rectangular matrix with more columns than rows. From Corollary A.6 we know that the columns of such a matrix must be linearly dependent; I.e. there is some nonzero $\boldsymbol{c} \in \mathbb{R}^k$ such that $\boldsymbol{A}\boldsymbol{c} = \boldsymbol{0}$. But then $\boldsymbol{\psi}^T \boldsymbol{c} = \boldsymbol{\phi}^T \boldsymbol{A}\boldsymbol{c} = \boldsymbol{0}$, for some nonzero $\boldsymbol{c}$. This implies that $\boldsymbol{\psi}$ is linearly dependent, a contradiction. We conclude that $k \le n$. ∎

Lemma A.18. *Every basis for a vector space must have the same number of elements.*

Proof. Suppose $\boldsymbol{\phi} = (\phi_1, \ldots, \phi_n)^T$ and $\boldsymbol{\psi} = (\psi_1, \ldots, \psi_m)^T$ are two bases for the vector space. We need to show that $m = n$. Now

$$\boldsymbol{\phi} = \boldsymbol{A}^T \boldsymbol{\psi}, \quad \text{for some} \quad \boldsymbol{A} \in \mathbb{R}^{m,n},$$

$$\boldsymbol{\psi} = \boldsymbol{B}^T \boldsymbol{\phi}, \quad \text{for some} \quad \boldsymbol{B} \in \mathbb{R}^{n,m}.$$

By Lemma A.16 we know that both $\boldsymbol{A}$ and $\boldsymbol{B}$ have linearly independent columns. But then by Corollary A.6 we see that $m = n$. ∎

Definition A.19. *The number of elements in a basis in a vector space S is called the* dimension *of S, and is denoted by dim(S).*

The following lemma shows that every set of linearly independent functions in a vector space S can be extended to a basis for S. In particular every finite dimensional vector space has a basis.

Lemma A.20. *A set $\boldsymbol{\phi}^T = (\phi_1, \dots, \phi_k)$ of linearly independent elements in a finite dimensional vector space S, can be extended to a basis $\boldsymbol{\psi}^T = (\psi_1, \dots, \psi_m)$ for S.*

Proof. Let $S_k = \mathrm{span}(\psi_1, \dots, \psi_k)$ where $\psi_j = \phi_j$ for $j = 1, \dots, k$. If $S_k = S$ then we set $m = k$ and stop. Otherwise there must be an element $\psi_{k+1} \in S$ such that $\psi_1, \dots, \psi_{k+1}$ are linearly independent. We define a new vector space S_{k+1} by $S_{k+1} = \mathrm{span}(\psi_1, \dots, \psi_{k+1})$. If $S_{k+1} = S$ then we set $m = k+1$ and stop the process. Otherwise we continue to generate vector spaces $S_{k+2}, S_{k+3}, \cdots$. Since S is finitely generated we must by Lemma A.17 eventually find some m such that $S_m = S$. ∎

The following simple, but useful lemma, shows that a spanning set must be a basis if it contains the correct number of elements.

Lemma A.21. *Suppose $S = \mathrm{span}(\boldsymbol{\phi})$. If $\boldsymbol{\phi}$ contains $\dim(S)$ elements then $\boldsymbol{\phi}$ is a basis for S.*

Proof. Let $n = \dim(S)$ and suppose $\boldsymbol{\phi} = (\phi_1, \dots, \phi_n)$ is a linearly dependent set. Then there is one element, say ϕ_n which can be written as a linear combination of $\phi_1, \dots, \phi_{n-1}$. But then $S = \mathrm{span}(\phi_1, \dots, \phi_{n-1})$ and $\dim(S) < n$ by Lemma A.17, a contradiction to the assumption that $\boldsymbol{\phi}$ is linearly dependent. ∎

A.4 Normed Vector Spaces

Suppose S is a vector space of functions. A *norm* $||\ || = ||f||$, is a function $\|\ \| : S \to [0, \infty)$ that satisfies for $f, g, \in S$, and $\alpha \in \mathbb{R}$ the following properties

$$\begin{array}{ll} 1. & ||f|| = 0 \quad \text{implies} \quad f = 0. \\ 2. & ||\alpha f|| = |\alpha| ||f||. \\ 3. & ||f + g|| \le ||f|| + ||g||. \end{array} \tag{A.13}$$

Property 3 is known as the *Triangle Inequality.* The pair $(S, ||\ ||)$ is called a normed vector space (of functions).

In the rest of this section we assume that the functions in S are continuous, or at least piecewise continuous on some interval $[a, b]$.

Analogous to the p or ℓ^p norms for vectors in $\mathbb{R}^n$ we have the p or L^p norms for functions. They are defined for $1 \le p \le \infty$ and $f \in S$ by

$$\begin{aligned} ||f||_p = ||f||_{L^p[a,b]} &= \left(\int_a^b |f(x)|^p dx\right)^{1/p}, \quad p \ge 1, \\ ||f||_\infty = ||f||_{L^\infty[a,b]} &= \max_{a \le x \le b} |f(x)|. \end{aligned} \tag{A.14}$$

The 1,2, and ∞ norms are the most important.

We have for $1 \le p \le \infty$ and $f, g \in S$ the Hölder inequality

$$\int_a^b |f(x)g(x)|dx \le ||f||_p ||g||_q, \quad \textit{where} \quad \frac{1}{p} + \frac{1}{q} = 1, \tag{A.15}$$

and the Minkowski inequality

$$||f+g||_p \leq ||f||_p + ||g||_p. \tag{A.16}$$

For $p = 2$ (A.15) is known as the Schwarz inequality, the Cauchy-Schwarz inequality, or the Buniakowski-Cauchy- Schwarz inequality.