Appendix A

Linear Algebra and Analysis

This short review is not intended as an introduction or tutorial. On the contrary, it is assumed that the reader is already familiar with multi-variable calculus, the basic facts concerning metric spaces, and the elementary theory of finite-dimensional real and complex vector spaces. The goal of this appendix is rather to clarify just what material is assumed, to develop a consistent notation and point of view towards these subjects for use in the rest of the book, and to formulate some of their concepts and propositions in ways that will be convenient for applications elsewhere in the text.

A.1. Metric and Normed Spaces

A metric space is just a set X with a *distance* $\rho(x_1, x_2)$ defined between any two of its points x_1 and x_2. The distance should be a non-negative real number that is zero if and only if $x_1 = x_2$, and it should be symmetric in x_1 and x_2. Aside from these obvious properties of anything we would consider calling a distance function, the only other property we demand of the function ρ (which is also called the *metric* of X) is that the "triangle inequality" hold for any three points x_1, x_2, and x_3 of X. This just means that $\rho(x_1, x_3) \leq \rho(x_1, x_2) + \rho(x_2, x_3)$, and what it says in words is that "things close to the same thing are close to each other".

If $\{x_n\}$ is a sequence of points in X, then we say this sequence converges to a point x in X if $\lim_{n\to\infty} \rho(x_n, x) = 0$. It follows from the triangle inequality that if the sequence also converges to x', then

$\rho(x, x') = 0$, so $x = x'$, i.e., the limit of $\{x_n\}$ is unique if it exists, and we write $\lim_{n\to\infty} x_n = x$. The sequence $\{x_n\}$ is called a *Cauchy sequence* if $\rho(x_n, x_m)$ converges to zero as both m and n tend to infinity. It is easy to check that a convergent sequence is Cauchy, but the reverse need not be true, and if every Cauchy sequence in X does in fact converge, then we call X a *complete* metric space.

If X and Y are metric spaces and $f : X \to Y$ is a function, then we call f continuous if and only if $f(x_n)$ converges to $f(x)$ whenever x_n converges to x. An equivalent definition is that given any positive ϵ and an x in X, there exists a $\delta > 0$ such that if $\rho_X(x, x') < \delta$, then $\rho_Y(f(x), f(x')) < \epsilon$, and if we can choose this δ independent of x, then we call f *uniformly continuous*. A positive constant K is called a Lipschitz constant for f if for all x_1, x_2 in X, $\rho_Y(f(x_1), f(x_2)) < K\rho_X(x_1, x_2)$, and we call f a *contraction* if it has a Lipschitz constant K satisfying $K < 1$. Note that if f has a Lipschitz constant (in particular, if it is a contraction), then f is automatically uniformly continuous (take $\delta = \epsilon/K$).

▷ **Exercise A–1.** Show that if K is a Lipschitz constant for $f : X \to Y$ and L is a Lipschitz constant for $g : Y \to Z$, then KL is a Lipschitz constant for $g \circ f : X \to Z$.

The classic example of a metric space is $\mathbf{R}$ with $\rho(x, y) = |x - y|$, and this has an important generalization. Namely, if V is any (real or complex) vector space, then a nonnegative real-valued function $v \mapsto \|v\|$ on V is called a *norm* for V if it shares three basic properties of the absolute value of a real number, namely i) positve homogeneity, i.e., $\|\alpha v\| = |\alpha| \, \|v\|$ for a scalar α; ii) $\|v\| = 0$ only if $v = 0$; and iii) the triangle inequality $\|v_1 + v_2\| \leq \|v_1\| + \|v_2\|$ for all $v_1, v_2 \in V$. A vector space V together with a choice of norm for V is called a normed vector space, and on such a V we get a metric by defining $\rho(v_1, v_2) = \|v_1 - v_2\|$. If this makes V a complete metric space, then the normed space V is called a Banach space. In particular, $\mathbf{R}^n$ and $\mathbf{C}^n$ with their usual norms are Banach spaces.

If A is a subset of a metric space X, then the metric for X restricted to $A \times A$ defines a metric for A, and the resulting metric space is called a subspace of X.

If x is a point of the metric space X and $\epsilon > 0$, then we denote $\{y \in X \mid \rho(x, y) < \epsilon\}$, the so-called "open ball of radius ϵ about x in X", by $B_\epsilon(x, X)$. A subset A of X is a neighborhood of x if it includes $B_\epsilon(x, X)$ for some positive ϵ, and A is called an open subset of X if it is a neighborhood of each of its points. On the other hand A is called a closed subset of X if the limit of any sequence of points in A is itself in A. It is easily proved that A is closed in X if and only if its complement in X is open and that $f : X \to Y$ is continuous if and only if the inverse image, $f^{-1}(O)$, of every open subset O of Y is open in X.

If X is a metric space, then X itself is both open and closed in X, and hence the same is true of its complement, the empty set. If there are no other subsets of X that are both open and closed (or equivalently if X cannot be partitioned into two complementary nonempty open sets), then X is called a *connected* space. We say that a subset A of X is connected if the corresponding subspace is a connected metric space. There is an important characterization of the connected subsets of $\mathbf{R}$, namely $A \subseteq \mathbf{R}$ is connected if and only if, it is an interval, i.e., if and only if, whenever it contains two points, it also contains all points in between.

A subset A of a metric space is called *compact* if every sequence in A has a subsequence that converges to a point of A. The Bolzano-Weierstrass Theorem characterizes the compact subsets of $\mathbf{R}^n$ (and $\mathbf{C}^n$). Namely they are precisely those sets that are both closed and bounded. (A subset A of a metric space is bounded if and only if the distances between points of A are bounded above.)

A.2. Inner-Product Spaces

An *inner product* on a real vector space V is a real-valued function on $V \times V$, $(x, y) \mapsto \langle x, y \rangle$ having the following three properties:

1) Symmetry: $\langle x, y \rangle = \langle y, x \rangle$ for all $x, y \in V$.
2) Positive definiteness: $\langle x, x \rangle \geq 0$, with equality if and only if $x = 0$.
3) Bilinearity: $\langle \alpha x + \beta y, z \rangle = \alpha \langle x, z \rangle + \beta \langle y, z \rangle$, for all $x, y, z \in V$ and all $\alpha, \beta \in \mathbf{R}$.

An *inner-product space* is a pair $(V, \langle\ ,\ \rangle)$ consisting of a real vector space V and a choice of inner product for V, but it is customary to suppress reference to the inner product. The motivating example of an inner-product space is of course $\mathbf{R}^n$ with the usual "dot-product" $\langle x, y\rangle := \sum_{i=1}^n x_i y_i$.

In what follows, V will denote an arbitrary inner-product space, and we define $\|x\|$, the norm of an element x of V, by $\|x\| := \sqrt{\langle x, x\rangle}$.

By bilinearity, if $x, y \in V$ and $t \in \mathbf{R}$, then $\|tx + y\|^2$ is a quadratic polynomial function of t, namely,

$$\|tx + y\|^2 = \langle tx + y, tx + y\rangle = \|x\|^2 t^2 + 2\langle x, y\rangle t + \|y\|^2,$$

and note the important special case

$$\|x + y\|^2 = \|x\|^2 + 2\langle x, y\rangle + \|y\|^2.$$

Finally, for reasons we shall see a little later, the two vectors x and y are called *orthogonal* if $\langle x, y\rangle = 0$, so in this case we have

A.2.1. Pythagorean Identity. *If x and y are orthogonal vectors in an inner product space, then $\|x + y\|^2 = \|x\|^2 + \|y\|^2$.*

Recall some basic high-school mathematics concerning a quadratic polynomial $P(t) = at^2 + bt + c$. (For simplicity, we assume a is positive.) The *discriminant* of $P(t)$ is the quantity $b^2 - 4ac$, and it distinguishes what kind of roots the polynomial has. In fact, the so-called "quadratic formula" says that the two (possibly complex) roots of $P(t)$ are $(-b \pm \sqrt{b^2 - 4ac})/2a$. So there are three cases:

Case 1: $b^2 - 4ac > 0$. Then $P(t)$ has two real roots. Between these roots $P(t)$ is negative, and outside of the interval between the roots it is positive.

Case 2: $b^2 - 4ac = 0$. Then $P(t)$ has only the single real root $-b/2a$, and elsewhere $P(t) > 0$.

Case 3: $b^2 - 4ac < 0$. Then $P(t)$ has no real roots, and $P(t)$ is positive for all real t.

For the polynomial $\|tx + y\|^2$ we see that $a = \|x\|^2$, $c = \|y\|^2$, and $b = 2\langle x, y\rangle$, so the discriminant is $4(|\langle x, y\rangle|^2 - \|x\|^2 \|y\|^2)$. Case 1 is ruled out by positive definiteness. In Case 2, we have $|\langle x, y\rangle| = \|x\|\,\|y\|$, so if t is the root of the polynomial, then $\|x + ty\| = 0$, so $x = -ty$, and we see that in this case x and y are linearly dependent. Finally, in Case 3, $|\langle x, y\rangle| < \|x\|\,\|y\|$, and since $x + ty$ is never zero, x and y are linearly independent. This proves one of the most important inequalities in all of mathematics.

A.2.2. Schwartz Inequality. *For all $x, y \in V$, $|\langle x, y\rangle| \leq \|x\|\,\|y\|$, with equality if and only if x and y are linearly dependent.*

▷ **Exercise A–2.** Use the Schwartz Inequality to deduce the triangle inequality:

$$\|x + y\| \leq \|x\| + \|y\| .$$

(Hint: Square both sides.)

This shows that an inner-product space is a normed space.

• **Example A–1.** Let $C([a, b])$ denote the vector space of continuous real-valued functions on the interval $[a, b]$. For $f, g \in C([a, b])$ define $\langle f, g\rangle = \int_a^b f(x)g(x)\,dx$. It is easy to check that this satisfies our three conditions for an inner product.

In what follows, we assume that V is an inner-product space. If $v \in V$ is a nonzero vector, we define a unit vector e with the same direction as V by $e := v/\|v\|$. This is called *normalizing* v, and if v already has unit length, then we say that v is *normalized*. We say that k vectors $e_1, \ldots, e_k$ in V are *orthonormal* if each e_i is normalized and if the e_i are mutually orthogonal. Note that these conditions can be written succinctly as $\langle e_i, e_j\rangle = \delta^i_j$, where δ^i_j is the so-called Kronecker delta symbol and is defined to be zero if i and j are different and 1 if they are equal.

▷ **Exercise A–3.** Show that if $e_1, \ldots, e_k$ are orthonormal and v is a linear combination of the e_i, say $v = \alpha_1 v_1 + \cdots + \alpha_k v_k$, then the α_i are uniquely determined by the formulas $\alpha_i = \langle v, e_i\rangle$. Deduce from this that orthonormal vectors are automatically linearly independent.

Orthonormal bases are also referred to as *frames* and they play an very important role in all explicit computation in inner-product spaces. Note that if $e_1, \ldots, e_n$ is an orthonormal basis for V, then every element of V is a linear combination of the e_i, so that by the exercise each $v \in V$ has the expansion $v = \sum_{i=1}^{n} \langle v, e_i \rangle \, e_i$.

• **Example A–2.** The "standard basis" for $\mathbf{R}^n$ is $\delta^1, \ldots, \delta^n$, where $\delta^i = (\delta_1^1, \ldots, \delta_n^i)$. It is clearly orthonormal.

Let V be an inner product space and W a linear subspace of V. We recall that the *orthogonal complement* of W, denoted by $W^\perp$, is the set of those v in V that are orthogonal to every w in W.

▷ **Exercise A–4.** Show that $W^\perp$ is a linear subspace of V and that $W \cap W^\perp = 0$.

If $v \in V$, we will say that a vector w in W is its orthogonal projection on W if $u = v - w$ is in $W^\perp$.

▷ **Exercise A–5.** Show that there can be at most one such w. (Hint: If w' is another, so $u' = v - u \in W^\perp$, then $u - u' = w' - w$ is in both W and $W^\perp$.)

A.2.3. Remark. Suppose $\omega \in W$. Then since $v - \omega = (v - w) + (w - \omega)$ and $v - w \in W^\perp$ while $(w - \omega) \in W$, it follows from the Pythagorean identity that $\|v - \omega\|^2 = \|v - w\|^2 + \|w - \omega\|^2$. Thus, $\|v - \omega\|$ is strictly greater than $\|v - w\|$ unless $\omega = w$. In other words, **the orthogonal projection of v on w is the unique point of W that has minimum distance from v.**

We call a map $P : V \to W$ the *orthogonal projection of V onto W* if $v - Pv$ is in $W^\perp$ for all $v \in V$. By the previous exercise this mapping is uniquely determined if it exists (and we will see below that it always does exist).

▷ **Exercise A–6.** Show that if $P : V \to W$ is the orthogonal projection onto W, then P is a linear map. Show also that if $v \in W$, then $Pv = v$ and hence $P^2 = P$.

▷ **Exercise A–7.** Show that if $e_1, \ldots, e_n$ is an orthonormal basis for W and if for each $v \in V$ we define $Pv := \sum_{i=1}^{n} \langle v, e_i \rangle \, e_i$, then P is orthogonal projection onto W. In particular, orthogonal projection onto W exists for any subspace W of V that has some orthonormal basis. Since, as we now will show, any W has an orthonormal basis, orthogonal projection on a subspace is always defined.

There is a beautiful algorithm, called the **Gram-Schmidt Procedure**, for starting with an arbitrary sequence $w_1, w_2, \ldots, w_k$ of linearly independent vectors in an inner-product space V and manufacturing an orthonormal sequence $e_1, \ldots, e_k$ out of them. Moreover it has the nice property that for all $j \leq k$, the sequence $e_1, \ldots, e_j$ spans the same subspace W_j of V as is spanned by $w_1, \ldots, w_j$.

In case $k = 1$ this is easy. To say that w_1 is linearly independent just means that it is nonzero, and we take e_1 to be its normalization: $e_1 := w_1 / \|w_1\|$. Surprisingly, this trivial special case is the crucial first step in an inductive procedure.

In fact, suppose that we have constructed orthonormal vectors $e_1, \ldots, e_m$ (where $m < k$) and that they span the same subspace W_m that is spanned by $w_1, \ldots, w_m$. How can we take the next step and construct e_{m+1} so that $e_1, \ldots, e_{m+1}$ is orthonormal and spans the same subspace as $w_1, \ldots, w_{m+1}$?

First note that since the $e_1, \ldots, e_m$ are linearly independent and span W_m, they are an orthonormal basis for W_m, and hence we can find the orthogonal projection ω_{m+1} of w_{m+1} onto W_m using the formula $\omega_{m+1} = \sum_{i=1}^{m} \langle w_{m+1}, e_i \rangle e_i$. Recall that this means that $\epsilon_{m+1} = w_{m+1} - \omega_{m+1}$ is orthogonal to W_m, and in particular to $e_1, \ldots, e_m$. Now ϵ_{m+1} **cannot be zero!** Why? Because if it were, then we would have $w_{m+1} = \omega_{m+1} \in W_m$, so w_{m+1} would be a linear combination of $w_1, \ldots, w_m$, contradicting the assumption that $w_1, \ldots, w_k$ are linearly independent. But then we can define e_{m+1} to be the normalization of ϵ_{m+1}, i.e., $e_{m+1} := \epsilon_{m+1} / \|\epsilon_{m+1}\|$, and it follows that e_{m+1} is also orthogonal to $e_1, \ldots, e_m$, so that $e_1, \ldots, e_{m+1}$ is orthonormal. Finally, it is immediate from its definition that e_{m+1} is a linear combination of $e_1, \ldots, e_m$ and w_{m+1} and hence of $w_1, \ldots, w_{m+1}$, completing the induction. Let's write the first few steps in the Gram-Schmidt Process explicitly.

(1)	$e_1 := w_1 / \|w_1\|$	% normalize w_1 to get e_1.
(2a)	$\omega_2 := \langle w_2, e_1 \rangle e_1$	% get projection ω_2 of w_2 on W_1.
(2b)	$\epsilon_2 := w_2 - \omega_2$	% subtract ω_2 from w_2 to get $W_1^{\perp}$ component ϵ_2 of w_2.

(2c) $e_2 := \epsilon_2 / \|\epsilon_2\|$ % and normalize it to get e_2.

(3a) $\omega_3 := \langle w_3, e_1 \rangle e_1 + \langle w_3, e_2 \rangle e_2$ % get projection ω_3 of w_3 on W_2.

(3b) $\epsilon_3 := w_3 - \omega_3$ % subtract ω_3 from w_3 to get $W_2^{\perp}$ component ϵ_3 of w_3.

(3c) $e_3 := \epsilon_3 / \|\epsilon_3\|$ % and normalize it to get e_3.

...

If W is a k-dimensional subspace of an n-dimensional inner-product space V, then we can start with a basis for W and extend it to a basis for V. If we now apply the Gram-Schmidt Procedure to this basis, we end up with an orthonormal basis for V with the first k elements in W and with the remaining $n - k$ in $W^{\perp}$. This tells us several things:

- $W^{\perp}$ has dimension $n - k$.
- V is the direct sum of W and $W^{\perp}$. This just means that every element of V can be written uniquely as the sum $w + u$ where $w \in W$ and $u \in W^{\perp}$.
- $(W^{\perp})^{\perp} = W$.
- If P is the orthogonal projection of V on W and I denotes the identity map of V, then $I - P$ is orthogonal projection of V on $W^{\perp}$.