Lecture 3
Geometry of Inner Product Spaces

3.1 Angles

Let x and y be two non-zero vectors in an inner product space. Then by the Schwartz inequality, the ratio $\langle x, y\rangle / \|x\| \|y\|$ lies in the interval $[-1, 1]$, so there is a unique angle θ between 0 and π such that $\cos(\theta) = \langle x, y\rangle / \|x\| \|y\|$. In other words, we define θ to make the identity $\langle x, y\rangle = \|x\| \|y\| \cos(\theta)$ hold. What is the geometric meaning of θ? Let's first consider a special case. Namely take for x the unit vector in the x direction, $(1, 0)$, and let y be an arbitrary vector in $\mathbf{R}^2$. If $r = \|y\|$ and ϕ is the angle between x and y (the so-called polar angle of y), then clearly $y = (r\cos(\phi), r\sin(\phi))$, so it follows that $\langle x, y\rangle = (1)(r\cos(\phi)) + (0)(r\sin(\phi)) = r\cos(\phi)$ and hence $\langle x, y\rangle / \|x\| \|y\| = \cos(\phi)$, so in this case the angle θ is exactly the angle ϕ between x and y.

▹ **3.1—Exercise 1.** Carry out the computation for the general case of two non-zero vectors in the plane with lengths r_1 and r_2 and polar angles ϕ_1 and ϕ_2, so that $x = (r_1\cos(\phi_1), r_1\sin(\phi_1))$ and $y = (r_2\cos(\phi_2), r_2\sin(\phi_2))$. Show that in this case too the ratio $\langle x, y\rangle / \|x\| \|y\|$ is the cosine of the angle $(\phi_1 - \phi_2)$ between x and y. (Hint: use the Cosine Addition Formula: $\cos(A \pm B) = \cos(A)\cos(B) \mp \sin(A)\sin(B)$.)

Henceforth we will refer to θ as the angle between x and y. In particular, if $\langle x, y\rangle = 0$, so that $\theta = \pi/2$, then we say that x and y are *orthogonal*.

3.2 Orthonormal Bases for an Inner Product Space

We begin by recalling the basic facts concerning linear dependence, dimension, and bases in a vector space V. (If you prefer to be concrete, you may think of V as being $\mathbf{R}^n$.) We say that vectors $v_1, \ldots, v_n$ in V are *linearly dependent* if there are scalars $\alpha_1, \ldots, \alpha_n$, **not all zero**. such that the linear combination $\alpha_1 v_1 + \cdots + \alpha_n v_n$ is the zero vector. It is easy to see that this is equivalent to one of the v_i being a linear combination of the others. If $v_1, \ldots, v_n$ are **not** linearly dependent, than we say that they are linearly independent. The vectors $v_1, \ldots, v_n$ are said to *span* V if every element of V can be written as a linear combination of the v_i, and if V is spanned by some finite set of vectors then we say that V finite dimensional, and we define the dimension of V, $\dim(V)$, to be the least number of vectors needed to span V. A finite set of vectors $v_1, \ldots, v_n$ in V is called a *basis* for V if it is both linearly independent and spans V. It is easy to see that this is equivalent to demanding that every element of V is a **unique** linear combination of the v_i. The following is a the basic theorem tying thse concepts together.

Theorem. *If V is an n-dimensional vector space, then every basis for V has exactly n elements. Moreover, if $v_1, \ldots, v_n$ is any set of n elements of V, then they form a basis for V if and only if they are linearly independent or if and only if they span V. In other words, n elements of V are linearly independent if and only if they span V.*

In what follows, we assume that V is an inner-product space. If $v \in V$ is a non-zero 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 $e_1, \dots, e_k$ 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.

▷ **3.2—Exercise 1.** Show that if $e_1, \dots, e_k$ 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 as we shall see they play an extremely important role in all things having to do with explicit computation in inner-product spaces. Note that if $e_1, \dots, e_n$ 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 $v = \sum_{i=1}^n \langle v, e_i \rangle e_i$.

3.2—Example 1. The "standard basis" for $\mathbf{R}^n$, is $\delta^1, \dots, \delta^n$, where $\delta^i = (\delta^1_1, \dots, \delta^i_n)$. It is clearly orthonormal.

3.3 Orthogonal Projection

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.

▷ **3.3—Exercise 1.** 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 $W^\perp$.

▷ **3.3—Exercise 2.** Show that there can be st 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 $W^\perp$.)

3.3.1 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 $P : V \to W$ *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).

▷ **3.3—Exercise 3.** Show that if $P : V \to W$ is orthogonal projection onto W, then P is a linear map. Show also that if $v \in W$, then $Pv = v$ and hence $P^2 = P$.

▷ **3.3—Exercise 4.** 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 the next section shows that any W has an orthonormal basis, orthogonal projction on a subspace is always defined.)

3.4 The Gram-Schmidt Algorithm

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 non-zero, 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 make 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 $w_1, \ldots, w_k$ were 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 for W and extend it to a basis for V. If we now apply Gram-Schmidt 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 $W^\perp$.

▷ **Project 1. Implement Gram-Schmidt as a Matlab Function**

In more detail, create a Matlab m-file GramSchmidt.m in which you define a Matlab function GramSchmidt(M) taking as input a rectangular matrix M of real numbers of arbitrary size $m \times n$, and assuming that the m rows of M are linearly independent, it should transform M into another $m \times n$ matrix in which the rows are orthonormal, and moreover such that the subspace spanned by the first k rows of the output matrix is the same as the space spanned by the first k rows of the input matrix. Clearly, in writting your algorithm, you will need to know the number of rows, m and the number of columns n of M. You can find these out using the Matlab size function. In fact, size(M) returns (m,n) while size(M,1) returns m and size(M,2) returns n. Your algorithm will have to do some sort of loop, iterating over each row in order. Be sure to test your function on a number of different matrices of various sizes. What happens to your function if you give it as input a matrix with linearly dependent rows. (Ideally it should report this fact and not just return garbage!)