Orthogonality and Projections

October 24, 2008

1 Introduction

Today we went pretty fast and covered some abstract topics. While you won't be tested directly on this material in your homework, understanding this lecture will help you in understanding linear algebra and perhaps in solving some "tricky" problems that appear on exams. If you're mentally swamped with change of basis and the Gram-Schmidt process then don't worry about this document. Otherwise, read on. If you are reading on, I strongly advise you to work out these problems with pen and paper in addition to reading. The transposes and dot-products can make your eyes glaze over pretty fast if you haven't actually worked through them by hand first...

Somehow everything we talk about stems from just three stupid little facts.

1) ${A^T}^T = A$ for any $m \times n$ matrix A
2) $(AB)^T = B^T A^T$ for any $m \times n$ matrix A and $n \times p$ matrix B
3) $[-\vec{v}^T-]\vec{y} = \vec{v} \cdot \vec{y}$ for any two vectors $\vec{v}, \vec{y} \in \mathbb{R}^n$.

Proposition 1.1. *Let A be a $m \times n$ matrix. Let $\vec{x} \in \mathbb{R}^n$ and $\vec{y} \in \mathbb{R}^m$ (notice $\vec{x}, \vec{y}$ live in different dimensions!). Then $A\vec{x} \cdot \vec{y} = \vec{x} \cdot A^T \vec{y}$*

This is true because

$$A\vec{x} \cdot \vec{y} = (A\vec{x})^T \vec{y} = (\vec{x}^T A^T)\vec{y} = \vec{x}^T (A^T \vec{y}) = \vec{x} \cdot A^T \vec{y}.$$

Proof. From left to right the reasons for equality are 3), 2), associativity of matrix multiplication, and 3) again. □

Next we prove a neat, very useful fact we'll need for later.

Proposition 1.2. *Suppose A is an $m \times n$ matrix with trivial nullspace (or equivalently, its columns are linearly independent). Then $A^T A$ is an invertible matrix.*

Proof. Because $A^T A$ is an $n \times n$ matrix, we need only show that $A^T A$ has trivial nullspace. To do this we must check that if $\vec{x} \in \mathbb{R}^n$ is nonzero, then $A^T A\vec{x}$ is also nonzero. One way to show that $A^T A\vec{x}$ is nonzero is to find a vector which is not orthogonal to (b/c the zero is vector is orthogonal to every vector). It turns out that it is always not orthogonal to $\vec{x}$ itself because:

$$A^T A\vec{x} \cdot \vec{x} = A^T(A\vec{x}) \cdot \vec{x} = A\vec{x} \cdot A^{T^T}\vec{x} = A\vec{x} \cdot A\vec{x} = |A\vec{x}|^2 \neq 0.$$

The last (non)equality holds because by assumption $\vec{x} \neq \vec{0}$ and A itself has nontrivial nullspace, so $A\vec{x}$ must be nonzero. Hence it must have nonzero length. Okay, so $A^T A$ has trivial nullspace, and must be invertible. □

Now we define orthogonality and orthonormality. If you'll bear with me, you'll see there is some relationship between the shit we just did above and orthonormal basis.

Definition 1.3. *Two vectors $\vec{v}, \vec{w}$ are orthogonal if $\vec{v} \cdot \vec{w} = 0$*

So I guess we can add "orthogonal" to the list which now includes "normal" and "perpendicular"....they all mean the same thing.

Definition 1.4. *A basis $\beta = \{\vec{v}_1, ..., \vec{v}_k\}$ is orthogonal if all the basis elements are orthogonal to each other, i.e. $\vec{v}_i \cdot \vec{v}_j = 0$ for all $1 \leq i \neq j \leq k$.*

Even more special than an orthogonal basis is an orthonormal one.

Definition 1.5. *A basis $\beta = \{\vec{v}_1, ..., \vec{v}_k\}$ is orthonormal if it is orthogonal and all the basis elements have unit norm, i.e. $|\vec{v}_i| = 1$ for all $1 \leq i \leq k$.*

Now suppose C is a $k \times k$ change of basis matrix from β to standard-coordinate where $\beta = \{\vec{v}_1, ..., \vec{v}_k\}$ is an orthonormal basis. Then the i^{th} column of C is $\vec{v}_i$ and the j^{th} row of C^T is $\vec{v}_j^T$. Then by multiplication formula for matrices tells us that the ij^{th} entry in the matrix $C^T C$ is $\vec{v}_i \cdot \vec{v}_j$. Since β is orthonormal, the entries along the diagaonal, $\vec{v}_i \cdot \vec{v}_i$ are equal to 1 and the entries off the diagonal, $\vec{v}_i \cdot \vec{v}_j$ with $i \neq j$ are zero. Thus $C^T C$ is the identity matrix. So we have found out that

Proposition 1.6. *Let* $\beta = \{\vec{v}_1, ..., \vec{v}_k\}$ *be an orthonormal basis and let* C *be the matrix whose* i^{th} *column is the vector* $\vec{v}_i$*. Then* $C^{-1}C = I_k$ *or equivalently* $C^T = C^{-1}$*.*

This should seem pretty useful. After all, C is the change of basis matrix from β to standard coordinates. Instead of having to find the inverse of C the usual way, we just need to take its transpose. By the way, it should be clear by following the argument backwards that the converse is true: that if $C^T = C^{-1}$ then the columns of C are orthonormal.

These kinds of matrices are special, and we might wonder what kind of linear transformations they represent. It turns out that is pretty easy to answer. These matrices represent linear transformations which preserve the angles between vectors (and their lengths):

Proposition 1.7. *If* $T : \mathbb{R}^k \to \mathbb{R}^k$ *is a linear transformation such that* $T\vec{x} \cdot T\vec{y} = \vec{x} \cdot \vec{y}$ *for all* $\vec{x}, \vec{y} \in \mathbb{R}^k$*, and if* A *is the matrix representing* T*, then* $A^T A = I_k$*. Conversely, if* $A^T A = I_k$*, then* $A\vec{x} \cdot A\vec{y} = \vec{x} \cdot \vec{y}$ *for all* $\vec{x}, \vec{y} \in \mathbb{R}^k$*.*

Proof. The converse is easy. If $A^T A = I_k$, then

$$A\vec{x} \cdot A\vec{x} = \vec{x} \cdot A^T A\vec{x} = \vec{x} \cdot I_k\vec{x} = \vec{x} \cdot \vec{x}.$$

Now lets prove the first statement. If A represents T, we just need to show that the columns of A are orthonormal. Then by our previous work, we'd know that $A^T A = I_k$. Remember that if A represents T, then the columns of A are, in order: $\{T\vec{e}_1, ..., T\vec{e}_k\}$. But since the standard basis vectors $\{\vec{e}_1, ..., \vec{e}_k\}$ are orthonormal, and by assumption T preserves dot products, it quickly follows that $\{T\vec{e}_1, ..., T\vec{e}_k\}$ is orthonormal. □

2 Projection

Okay, in this section we're going to show how to define projection fairly rigorously. Now some of you may object and say that we already know how, but I say that as of now, we only knew how to project onto a line. Here is the real deal. Let me try to summarize first.

The idea is that if $V \subset \mathbb{R}^n$ is a k-dimensional subspace, then there is a space of vectors normal to V which we denote by $V^\perp$ which is $(n-k)$-dimensional. Formally $V^\perp$ is defined as

$$V^\perp := \{ \vec{x} \in \mathbb{R}^n \mid \vec{x} \cdot \vec{v} = 0 \text{ for all } \vec{v} \in V \}.$$

(Aside: you should try on your own to prove that $V^\perp$ is a subspace. Also in class we found $V^\perp$ was when $V = \text{span}\{\begin{bmatrix} 1 \\ 2 \\ 0 \\ -3 \end{bmatrix}\}$. MAKE SURE this is second nature to you. If not, come see me. End Aside)

The two spaces V and $V^\perp$ are special in that

Any vector $\vec{x} \in \mathbb{R}^n$ may be decomposed **uniquely** as a sum of two vectors,

$$\vec{x} = \vec{x}_V + \vec{x}_{V^\perp} \text{ where } \vec{x}_V \in V \text{ and } \vec{x}_{V^\perp} \in V^\perp.$$

The vector $\vec{x}_V$ is what we call $\text{Proj}_V \vec{x}$ and the vector $\vec{x}_{V^\perp}$ is what we call $\text{Proj}_{V^\perp} \vec{x}$. The formula

$$\vec{x} = \vec{x}_V + \vec{x}_{V^\perp}$$

can be rewritten as

$$I_n = \text{Proj}_V + \text{Proj}_{V^\perp}.$$

This formula is useful. If we somehow find the matrix representing Proj_V, then it is just a matter of subtraction to find $\text{Proj}_{V^\perp}$.

We will also see that there is a formula for finding the matrix which represents Proj_V which we can figure out once we have selected a basis for V. If the basis happens to be orthonormal, then the formula takes a simple form. OK, enough summarizing. We need to start by proving the important statement

Proposition 2.1.

*Any vector $\vec{x} \in \mathbb{R}^n$ may be decomposed **uniquely** as a sum of two vectors,*

$$\vec{x} = \vec{x}_V + \vec{x}_{V^\perp} \textit{ where } \vec{x}_V \in V \textit{ and } \vec{x}_{V^\perp} \in V^\perp.$$

Proof. The proposition follows if we can prove two facts:

1) $V \cap V^\perp = \{\vec{0}\}$
2) $V + V^\perp = \mathbb{R}^n$.

Lets see why these two facts imply the main statement. The second statement says that any vector $\vec{x}$ can be written uniquely as a sum of two vectors, one in V, the other in $V^\perp$. It remains to show that there is a *unique* way to do this. Lets first show that there is a unique way to decompose the zero vector in this way (as $\vec{0}+\vec{0}$, duh). Suppose we could do it another way, $\vec{0} = \vec{x}_V + \vec{x}_{V^\perp}$ where the vectors on the RHS are nonzero. Well, then subtracting one to the other side, we get $\vec{x}_V = -\vec{x}_{V^\perp}$. As $V^\perp$ is a subspace, $-\vec{x}_{V^\perp}$ must also lie in $V^\perp$. Then the vector $\vec{x}_V$ lies in both V and $V^\perp$. By fact 1), this means $\vec{x}_V = \vec{0}$. Thus there is only one way to decompose $\vec{0}$ as the sum of two vectors in V and $V^\perp$. Now lets show this is true for any vector $\vec{x}$ in $\mathbb{R}^n$. Suppose $\vec{x}$ can be decomposed in two distinct ways:

$$\vec{x}_V + \vec{x}_{V^\perp} = \vec{x} = \vec{x}'_V + \vec{x}'_{V^\perp}.$$

Subtracting one side entirely and regrouping by subspaces we get

$$(\vec{x}_V - \vec{x}'_V) + (\vec{x}_{V^\perp} - \vec{x}'_{V^\perp}) = \vec{0}.$$

V and $V^\perp$ are subspaces, so by closure under addition, the vector $\vec{x}_V - \vec{x}'_V$ lies in V and $\vec{x}_{V^\perp} - \vec{x}_{V^\perp}$ lies in $V^\perp$. But now we're in the previous situation, where we've decomposed $\vec{0}$ as the sum of two vectors one in V and another in $V^\perp$. We know then that both vectors must be zero. Thus

$$\vec{x}_V - \vec{x}'_V = \vec{0} \text{ and } \vec{x}_{V^\perp} - \vec{x}_{V^\perp} = \vec{0}$$

so $\vec{x}_V = \vec{x}'_V$ and $\vec{x}_{V^\perp} = \vec{x}'_{V^\perp}$ so evidently the two decompositions were actually the same. Whew. Okay now it remains to verify the two facts 1) and 2). 1) is easy, 2) is harder and uses that A^TA crap we did at the beginning.

1) Suppose $\vec{x} \in V \cap V^\perp$. Then by definition of being in $V^\perp$, $\vec{x}$ must be normal to all vectors in V. Since $\vec{x}$ is in V too, it must be normal to itself, i.e. $\vec{x} \cdot \vec{x} = 0$. The only vector with this property is the zero vector. So $V \cap V^\perp = \{\vec{0}\}$.

2) Okay, this one is long and hard, I know, but you'll be better for it...that's a lie...we'll all be just as pathetic as we were before going over it. Maybe even sadder. But anyway... We need to show that $\mathbb{R}^n = V + V^\perp$. The strategy is to show that $V^\perp$ is $(n-k)$-dimensional (remember, V was assumed to be k-dimensional). Then we will show that if we take a basis $\{\vec{v}_1, ..., \vec{v}_k\}$ for V and a basis $\{\vec{w}_1, ..., \vec{w}_{n-k}\}$ for $V^\perp$ and join them together to get a mega-collection $\{\vec{v}_1, ..., \vec{v}_k, \vec{w}_1, ..., \vec{w}_{n-k}\}$ then this mega-collection is still linearly independent. As it has n elements, it must be a basis for $\mathbb{R}^n$. On the otherhand any linear combination of these vectors is obviously in $V + V^\perp$ thus $V + V^\perp = \mathbb{R}^n$. Okay, so it looks like we need to only show two

things then:

2a) dim$V^\perp = n - k$

2b) If $\{\vec{v}_1, ..., \vec{v}_k\}$ is a basis for V and $\{\vec{w}_1, ..., \vec{w}_{n-k}\}$ is a basis for $V^\perp$ then $\{\vec{v}_1, ..., \vec{v}_k, \vec{w}_1, ..., \vec{w}_{n-k}\}$ is linearly independent.

2a) Fix a basis $\{\vec{v}_1, ..., \vec{v}_k\}$ for V. Let A be the $n \times k$ matrix whose i^{th} column is the vector $\vec{v}_i$. Notice that $V^\perp = N(A^T)$. This is almost by definition, since a vector in the nullspace of A^T is a vector normal to all the vectors $\vec{v}_i$ (please, come see me if you can't see this by now). Thus it suffices to show that dim$N(A^T) = n - k$ or by the Rank-Nullity Theorem, that dim$C(A^T) = k$. This is what we will do. As is always the case in this class, when we want to show something like this we do it in two steps. We show that dim$C(A^T) \leq k$ and dim$C(A^T) \geq k$.

As A^T is a $k \times n$ matrix, $C(A)$ is a subspace of $\mathbb{R}^k$ thus dim$C(A^T) \leq k$. Now we must show the other inequality. To this end, notice that $C(A^T A)$ is a subspace of $C(A^T)$. Lets explain this, it is easy once you see it. $C(A^T)$ is equal to the set of vectors $A^T \vec{x}$ where $\vec{x}$ is allowed to be any vector in $\mathbb{R}^n$. $C(A^T A)$ on the otherhand is the set of vectors $A^T \vec{y}$ where $\vec{y}$ is any vector which is in $C(A)$. Since $C(A)$ is a subspace of $\mathbb{R}^n$, this set evidently includes in $C(A^T)$. Okay, well since $C(A^T A) \subset C(A^T)$ we must have dim$C(A^T A) \leq$ dim$C(A^T)$. But now we remember that $A^T A$ is an invertible $k \times k$ matrix when the columns of A are linearly independent (which is the case here). Since $A^T A$ is invertible, we must have that dim$C(A^T A) = k$, so $k \leq$ dim$C(A^T)$. This is what we needed to show. Whew. So $V^\perp$ is $(n - k)$ dimensional.

2b) We didn't have time to prove this in class. If you're feeling adventurous, try to do it without reading the version below!

We want to show $\{\vec{v}_1, ..., \vec{v}_k, \vec{w}_1, ..., \vec{w}_{n-k}\}$ is linearly independent. To that end, suppose

$$\vec{0} = c_1 \vec{v}_1 + \cdots + c_k \vec{v}_k + d_1 \vec{w}_1 + \cdots + d_{n-k} \vec{w}_{n-k}.$$

Subtract all the $c_i \vec{v}_i$'s to the other side:

$$-(c_1 \vec{v}_1 + \cdots + c_k \vec{v}_k) = d_1 \vec{w}_1 + \cdots + d_{n-k} \vec{w}_{n-k}.$$

The lefthand side is clearly an element of V while the righthand side is clearly an element of $V^{\perp}$. Since they are equal, this value must evidently lie in both V and $V^{\perp}$, i.e. it lies in $V \cap V^{\perp}$. But by 1) we found that the only element in $V \cap V^{\perp}$ is $\vec{0}$. So

$$\vec{0} = c_1\vec{v}_1 + \cdots + c_k\vec{v}_k \text{ and } \vec{0} = d_1\vec{w}_1 + \cdots + d_{n-k}\vec{w}_{n-k}.$$

Since the collection $\{\vec{v}_1, ..., \vec{v}_k\}$ is linearly independent we conclude that all the c_i's are zero. Similarly we conclude all the d_i's are zero. DONE! Oh wait, now where are we...oh yes.... □

3 Formula for calculating projection

Now that we know projections exist, here is the formula. Let V be subspace of $\mathbb{R}^n$. Let $\{\vec{v}_1, ..., \vec{v}_k\}$ be a basis for V and let A be the matrix whose i^{th} column is the vector $\vec{v}_i$. Then

$$\text{Proj}_V = A(A^T A)^{-1} A^T.$$

This is a weird looking formula, but it makes sense mathgrammatically. First off we found that $A^T A$ is invertible, so it makes sense to write $(A^T A)^{-1}$. By the way, why can't we use the fact that $(AB)^{-1} = B^{-1}A^{-1}$ to say $A(A^T A)^{-1} A = A(A^{-1} A^{T^{-1}}) A^T = (AA^{-1})(A^{T^{-1}} A^T) = (Id)(Id) = Id$? Send me an email when you see why.

Lets check that $A(A^T A)^{-1} A^T \vec{x}$ agrees with the abstract definition of $\text{Proj}_V \vec{x}$ we gave earlier. This means we must check that $A(A^T A)^{-1} A^T \vec{x}$ is none other that the vector $\vec{x}_V$, where remember, $\vec{x}_V \in V$ is that vector that sums with a vector in $V^{\perp}$ to $\vec{x}$. To check this, we only need to verify that $A(A^T A)^{-1} A^T \vec{x}$ is in V and that $\vec{x} - A(A^T A)^{-1} A^T \vec{x}$ lies in $V^{\perp}$. Well, the first follows since the last matrix we multiply by in $A(A^T A)^{-1} A^T \vec{x}$ is the matrix A whose columns are vectors in V. Then $A(A^T A)^{-1} A^T \vec{x} = A((A^T A)^{-1} A^T \vec{x})$ must be some linear combination of those vectors, which certainly lies in V. Now to check that $\vec{x} - A(A^T A)^{-1} A^T \vec{x}$ lies in $V^{\perp}$ we use the definition of $V^{\perp}$. This means we must check that $\vec{x} - A(A^T A)^{-1} A^T \vec{x}$ is normal to every vector in V. The trick here is that by construction, $C(A) = V$ (remember, the columns of A are the elements for a basis of V), so we may write any vector in V as $A\vec{y}$ for some $\vec{y} \in \mathbb{R}^n$. Now we want to see what happens when we dot $A\vec{y}$ by $\vec{x} - A(A^T A)^{-1} A^T \vec{x}$. In fact for now, lets just see what happens when we dot $A\vec{y}$ with $A(A^TA)^{-1}A^T\vec{x}$. We can probably expect to use the rule $A\vec{y}\cdot\vec{x} = \vec{y}\cdot A^T\vec{x}$...twice.

$$A\vec{y}\cdot A(A^TA)^{-1}A^T\vec{x} = \vec{y}\cdot A^TA(A^TA)^{-1}A^T\vec{x}.$$

That was the first time, and now we see buried in the expression $A^TA(A^TA)^{-1}$ which cancels, leaving us with

$$= \vec{y}\cdot A^T\vec{x}$$

now we use the transpose/dot-product rule once more yielding

$$= A\vec{y}\cdot\vec{x}.$$

Okay, well now we're ready.

$$A\vec{y}\cdot(\vec{x} - A(A^TA)^{-1}A^T\vec{x}) = A\vec{y}\cdot\vec{x} - A\vec{y}\cdot A(A^TA)^{-1}A^T\vec{x}$$

and now we substitute in our previous computation to find this is equal to

$$= A\vec{y}\cdot\vec{x} - A\vec{y}\cdot\vec{x} = 0.$$

So yes, $\vec{x} - A(A^TA)^{-1}A^T\vec{x}$ is normal to every element of V, so it is an element of $V^\perp$. Thus the vector $A(A^TA)^{-1}A^T\vec{x}$ fits our abstract critereon for being the projection of $\vec{x}$ on V. Okay, I promise that was the last long difficult problem.

4 What a pain in the ass and why would I want to use that formula anyway and this is stupid

So it isn't all that bad. Suppose the basis $\{\vec{v}_1, ..., \vec{v}_k\}$ for V that you chose was orthonormal. Then from the first part of section, we remember that the matrix A^TA is just the identity so the expression $A(A^TA)^{-1}A^T$ reduces to AA^T. If $\vec{x}\in\mathbb{R}^n$, then $AA^T\vec{x}$ reduces to a nice formula. $A^T\vec{x}$ is just the vector whose i^{th} component is $\vec{v}_i\cdot\vec{x}$. Then the vector $AA^T\vec{x}$ is given by the formula

$$AA^T\vec{x} = (\vec{v}_1\cdot\vec{x})\vec{v}_1 + \cdots + (\vec{v}_k\cdot\vec{x})\vec{v}_k.$$

So this is a nice projection formula. Not too bad I guess. Remember though, it only holds when you are using an orthonormal basis of V. ...well, we'll learn how to make orthonormal bases using the Gram-Schmidt process another time. I'm really tired now and its Thursday night, so you should be drinking away your pain with aftershave.