Inner Products

This learning module covers the following ideas. When you make your lesson plan, it should explain and contain examples of the following:

1. Explain the dot product and inner product. Use them to define length and angles of vectors. Normalize vectors, obtain vectors of a given length in a given direction, and explain how to tell if two vectors are orthogonal.

2. Obtain the orthogonal complement of a set of vectors (relate it to the null space of the transpose). Compute the cross product and project vectors onto other vectors.

3. By considering the projection of a vector onto a vector subspace, explain why solving $A^T A \vec{x} = A^T \vec{b}$ gives the solution to the least squares regression problem. This is the main objective of the unit, explaining why statistics works.

4. Use the Gram-Schmidt orthogonalization process to obtain an orthonormal basis of vectors. Explain why using this basis you obtain the simple equation $P^{-1} = P^T$ for the change of basis matrix. Show how to compute components of vectors relative to an orthogonal basis (called Fourier coefficients).

5. Illustrate by examples that when a matrix is symmetric, eigenvectors corresponding to different eigenvalues are orthogonal, which means you can find an orthonormal basis to diagonalize the matrix.

1 Preparation and Suggested Homework

	Preparation Problems
Day 1	7.2, 7.3, 7.3, 7.5
Day 2	7.6, 7.9, 7.10, 7.12
Day 3	7.14, 7.19, 7.20, 7.25 and 26
Day 4	7.22, 7.27, 11.25, 11.27

Homework list:

Concept	Sec.	Suggestions	Relevant Problems
Dot and Inner Product	7	1-5,7	1-7, 53-56
Orthogonality	7	9-14,	9-14, 60-67
Gram-Schmidt Process	7	18-23	18-23, 70-74
Orthogonal Matrices	7	24-28	24-28, 75-77
Diagonalizing Symmetric Matrices	11	25, 27	25-30, 67-70
More with Orthogonal Matrices	3		32-35, 97-100

Make sure you also review the linear regression problems from the second unit on Applications. In this unit, our main focus is to learn how inner products are used to give us lengths, angles, projections, and then linear regression. Along the way we will pick up some important vocabulary which is used in future generalizations of these ideas to other subjects. Orthonormal bases are extremely nice bases to use, since the corresponding matrix is orthogonal meaning $P^{-1} = P^T$ (inverting is simply transposing).

2 The dot and inner product

To find the length of a vector $\vec{u} = (u_1, u_2)$ in 2D, we use the Pythagorean theorem to find $||\vec{u}|| = |\vec{u}| = \sqrt{u_1^2 + u_2^2}$. The notation $|\vec{u}|$ and $||\vec{u}||$ are both used to denote the length or **norm** or a vector, depending on the context and author. In 3D, repeated use of the Pythagorean theorem gives $|(u_1, u_2, u_3)| = \sqrt{u_1^2 + u_2^2 + u_3^2}$. It seems reasonable to generalize this to $\mathbb{R}^n$ by letting $|\vec{u}| = |(u_1, \ldots, u_n)| = \sqrt{u_1^2 + \cdots u_n^2}$. Upon closer inspection, notice this is really just the matrix product $\sqrt{\vec{u}^T \vec{u}}$, for example we have

$$|(u_1, u_2)| = \sqrt{u_1^2 + u_2^2} = \sqrt{\begin{bmatrix} u_1 & u_2 \end{bmatrix} \begin{bmatrix} u_1 \\ u_2 \end{bmatrix}} \quad |(u_1, \ldots, u_n)| = \sqrt{u_1^2 + \cdots + u_n^2} = \sqrt{\vec{u}^T \vec{u}}$$

The above shows us that the square of the length of a vector is precisely the quantity $||\vec{u}||^2 = \vec{u}^T\vec{u}$. In trying to understand applications, mathematicians discovered that sometimes the quantity $\vec{u}^T\vec{v}$ appears where $\vec{u}$ and $\vec{v}$ are different vectors. This quantity appeared so often that they gave it a name, the dot product. (The word product is used because when taking the derivative of functions of vectors, the product rule applies to the dot product. Similarly, the matrix product satisfies the product rule, as does any mathematical "product.")

Given vectors $\vec{u} = (u_1, u_2, \ldots, u_n)$ and $\vec{v} = (v_1, v_2, \ldots, v_n)$, define the dot product $\vec{u} \cdot \vec{v}$ to be the scalar

$$\vec{u} \cdot \vec{v} = \vec{u}^T\vec{v} = u_1v_1 + u_2v_2 + \cdots u_nv_n = \sum_{i=1}^{n} u_iv_i.$$

If $\vec{u} = \vec{v}$, then the dot product $\vec{u} \cdot \vec{u} = ||\vec{u}||^2$ equals the length of the vector. If the length of a vector is 1, we say the vector is a **unit vector**. The standard basis vectors for $\mathbb{R}^n$ are unit vectors. Unit vectors give us a way to find vectors of any length in a given direction, all you have to do is multiply the unit vector by the length you want. In other words, every vector can always be written as a product of its length and a unit vector in its direction. If a vector is not a unit vector, we can make it a unit vector by dividing by its length. This process (obtaining a unit vector by dividing by length) is called **normalizing** a vector.

Example 1. Let's find a vector of length 7 in the direction of $(1, 3, -2)$. The vector $(1, 3, -2)$ has length $\sqrt{1+9+4} = \sqrt{14}$, so a unit vector in the same direction as $(1, 3, -2)$ is $\dfrac{1}{\sqrt{14}}(1, 3, -2) = \left(\dfrac{1}{\sqrt{14}}, \dfrac{3}{\sqrt{14}}, -\dfrac{2}{\sqrt{14}}\right)$. Since this vector is 1 long, multiply by 7 to obtain a vector which is 7 units long in the same direction. The vector we seek is $\dfrac{7}{\sqrt{14}}(1, 3, -2) = \left(\dfrac{7}{\sqrt{14}}, \dfrac{21}{\sqrt{14}}, -\dfrac{14}{\sqrt{14}}\right)$.

The dot product satisfies the following properties of arithmetic. The first three basically say it behaves like you would expect it to.

1. $(a\vec{u}) \cdot (b\vec{v}) = (ab)\vec{u} \cdot \vec{v}$ (You can pull out constants.)

2. $(\vec{u} + \vec{v}) \cdot \vec{w} = \vec{u} \cdot \vec{w} + \vec{v} \cdot \vec{w}$ (You can distribute.)

3. $\vec{u} \cdot \vec{v} = \vec{v} \cdot \vec{u}$ (It is commutative.)

4. $\vec{u} \cdot \vec{u} \geq 0$ and $\vec{u} \cdot \vec{u} = 0$ if and only if $\vec{u} = \vec{0}$.

We'll use these properties to show that the dot product provides a nice way to compute angles between vectors in 2D and 3D. Given two vectors $\vec{u}$ and $\vec{v}$ with their base at the origin, let θ be the angle between them. The two vectors form a triangle whose third edge is the vector $\vec{u} - \vec{v}$ (the vector connecting the heads of $\vec{u}$ and $\vec{v}$). Using the law of cosines from trigonometry ($c^2 = a^2 + b^2 - 2ab\cos\theta$), we have

$$|\vec{u} - \vec{v}|^2 = |\vec{u}|^2 + |\vec{v}|^2 - 2|\vec{u}||\vec{v}|\cos\theta.$$

Putting three lengths in terms of the dot product, we have $(\vec{u} - \vec{v}) \cdot (\vec{u} - \vec{v}) = \vec{u} \cdot \vec{u} + \vec{v} \cdot \vec{v} - 2|\vec{u}||\vec{v}|\cos\theta$. Expanding the left hand side we have $\vec{u} \cdot \vec{u} - 2\vec{u} \cdot \vec{v} + \vec{v} \cdot \vec{v} = \vec{u} \cdot \vec{u} + \vec{v} \cdot \vec{v} - 2|\vec{u}||\vec{v}|\cos\theta$ or canceling like terms on each side and dividing by -2 we have $\vec{u} \cdot \vec{v} = |\vec{u}||\vec{v}|\cos\theta$. This equation relates the dot product to the angle between two vectors, and we have the equations

$$\cos\theta = \frac{\vec{u} \cdot \vec{v}}{|\vec{u}||\vec{v}|} \qquad \theta = \cos^{-1}\left(\frac{\vec{u} \cdot \vec{v}}{|\vec{u}||\vec{v}|}\right).$$

The equations above work in 2D and 3D just the same, so in higher dimensions we will define the angle between two vectors to be the quantity above. In other words, the dot product gives us a way of talking about angles in all dimensions (why you would want to do such a thing is one purpose of this unit, out of which linear regression will emerge). In order to make this generalization, we need to know that $\frac{\vec{u} \cdot \vec{v}}{|\vec{u}||\vec{v}|} \leq 1$ so that it will always be the cosine of some angle. This is equivalent to showing $\vec{u} \cdot \vec{v} \leq |\vec{u}||\vec{v}|$, which is called the Cauchy-Schwartz inequality. Its proof is in the text.

(Draft March 25, 2009)

Example 2. The dot product of the two vectors $\vec{u} = (1, 0, 3, -4)$ and $\vec{v} = (3, -1, 5, 0)$ is $\vec{u} \cdot \vec{v} = (1)(3) + (0)(-2) + (3)(5) + (-4)(0) = 18$. The length of each vector is $|\vec{u}| = \sqrt{9 + 1 + 16} = \sqrt{26}$ and $|\vec{v}| = \sqrt{9 + 1 + 25} = \sqrt{35}$. The angle between $\vec{u}$ and $\vec{v}$ is $\theta = \cos^{-1}\left(\frac{18}{\sqrt{26}\sqrt{35}}\right)$ (which you can obtain in radians or degrees using a calculator).

The dot product of the two vectors $\vec{u} = (-2, 1, 3)$ and $\vec{v} = (1, 5, -1)$ is $-2 + 5 - 3 = 0$, so $\cos\theta = 0$ (without finding the lengths). This means that the angle between the vectors is 90 degrees, so the vectors are "perpendicular."

When the dot product of two vectors is zero, we say the vectors are orthogonal. Provided neither vector is zero, this means that the angle between the two vectors is 90 degrees. In summary, dot product helps us

1. Find the length of a vector: $\vec{u} \cdot \vec{u} = |\vec{u}|^2$,

2. Find the angle between two vectors: $\vec{u} \cdot \vec{v} = |\vec{u}||\vec{v}|\cos\theta$,

3. Determine if two vectors are orthogonal: $\vec{u} \cdot \vec{v} = 0$.

2.1   Inner Product - a generalization of the dot product

The dot product can be defined as the product $\vec{u}^T\vec{v}$. As mathematicians solved various problems, they discovered that if A is a symmetric matrix (meaning $A^T = A$) whose eigenvalues are positive, then $\vec{u}^T A\vec{v}$ provided a different kind of useful product. They then sought to define a new kind of product, called an inner product, by specifying some key properties the product must satisfy. These properties are essentially the same properties as the ones listed for the dot product.

Definition 1. An (real) inner product on a vector space V is a function $\langle \vec{u}, \vec{v} \rangle$ which satisfies the following 3 axioms:

1. $\langle a\vec{u} + b\vec{v}, \vec{w} \rangle = a \langle \vec{u}, \vec{w} \rangle + b \langle \vec{v}, \vec{w} \rangle$ (it's linear)

2. $\langle \vec{u}, \vec{v} \rangle = \langle \vec{v}, \vec{u} \rangle$ (it's symmetric)

3. $\langle \vec{u}, \vec{u} \rangle \geq 0$ and $\langle \vec{u}, \vec{0} \rangle = 0$ if and only if $\vec{u} = \vec{0}$. (it's positive definite)

A vector space together with an inner product is called an inner product space.

These three properties above imply many other properties, however the three properties provide in some sense a minimal set of assumptions to obtain something which behaves like the dot product. The dot product is called the usual inner product on $\mathbb{R}^n$, however there are many different kinds of inner products. We can repeat all the computations above to show that if $\langle \vec{u}, \vec{v} \rangle$ is an inner product on a vector space V, then relative to this inner product we can

1. Find the length or norm of a vector: $\langle \vec{u}, \vec{u} \rangle = ||\vec{u}||^2$,

2. Find the angle between two vectors :$\langle \vec{u}, \vec{v} \rangle = ||\vec{u}||||\vec{v}||\cos\theta$,

3. Determine if two vectors are orthogonal: $\langle \vec{u}, \vec{v} \rangle = 0$.

A proof that we can define angles in this way requires proving the Cauchy-Swartz inequality (pg 247), but we will just assume we can. The inner product is just a generalization of the dot product which allows us to work with vector spaces other than $\mathbb{R}^n$.

Example 3. The Federal Communications Commission (FCC) creates regulations about the types of signals that can be broadcast. Because of these regulations, signals form a vector space (a vector space of periodic function). If we consider the vector space of continuous functions on the interval $[0, 1]$, then we can define an inner product on this space by $\langle f, g \rangle = \int_0^1 fg\,dx$. We will now show that this is an inner product by showing it satisfies the three axioms.

1. $\langle af + bg, h \rangle = \int_0^1 (af + bg)h\,dx = \int_0^1 (afh + bgh)\,dx = a\int_0^1 fh\,dx + b\int_0^1 gh\,dx = a\langle f, h \rangle + b\langle g, h \rangle$ (so it's linear)

2. $\langle f, g \rangle = \int_0^1 fg \, dx = \int_0^1 gf \, dx = \langle g, f \rangle$ (so it's symmetric)

3. $\langle f, f \rangle = \int_0^1 f^2 dx \geq 0$ (the area under a curve which is always above the x-axis cannot be negative). If $f \neq 0$, then there will be some small interval on which it is positive, which means f^2 will have a nonzero amount of area under it. Hence $\langle f, f \rangle = 0$ if and only if $f = 0$ (so it's positive definite).

This example is essentially the foundational tool needed to fully understand how signals are transferred from radio towers to TV's, cell phone calls are transferred to and from communication satellites, etc.

Example 4. If A is a symmetric matrix which satisfies $\vec{x}^T A \vec{x} > 0$ for $\vec{x} \neq 0$ (which is equivalent to having real positive eigenvalues - this requires proof but for now we'll just use it as a fact), then an inner product on $\mathbb{R}^n$ can always be obtained by using the formula $\langle \vec{u}, \vec{v} \rangle = \vec{u}^T A \vec{v}$. Let's check the three conditions:

1. $\langle a\vec{u} + b\vec{v}, \vec{w} \rangle = (a\vec{u} + b\vec{v})A\vec{w} = a\vec{u}A\vec{w} + b\vec{v}A\vec{w} a \langle \vec{u}, \vec{w} \rangle + b \langle \vec{u}, \vec{w} \rangle$ (so it's linear)

2. The transpose of the number $\vec{v}^T A \vec{w}$ is $\vec{w}^T A^T \vec{v}$ (remember you reverse the order and transpose each matrix). Since A is symmetric, we have $\vec{w}^T A^T \vec{v} = \vec{w}^T A \vec{v}$. Since $\vec{v}^T A \vec{w}$ is a number, it equals its transpose which means $\vec{v}^T A \vec{w} = \vec{w}^T A \vec{v}$ or $\langle \vec{v}, \vec{w} \rangle = \langle \vec{w}, \vec{v} \rangle$.

3. We are assuming $\vec{x}^T A \vec{x} > 0$ for $\vec{x} \neq 0$ (or that the eigenvalues are positive), so it is immediately positive definite. We say a symmetric matrix is positive definite if it has the property (which is equivalent to being symmetric and having all positive eigenvalues).

The usual inner product is obtain by letting $A = I$, the identity matrix. If you want to show that something is an inner product, one approach is to show that it is represented by a symmetric matrix with positive eigenvalues. For example $\langle (u_1, u_2), (v_1, v_2) \rangle = 3u_1 v_1 - 1u_2 v_1 + 3u_2 v_2$ can be written in the form $\begin{bmatrix} u_1 & u_2 \end{bmatrix} \begin{bmatrix} 3 & -1 \\ -1 & 3 \end{bmatrix} \begin{bmatrix} v_1 \\ v_2 \end{bmatrix}$. The eigenvalues of this matrix satisfy $(3-\lambda)^2 - 1 = \lambda^2 - 6\lambda + 8 = (\lambda - 4)(\lambda - 2) = 0$, and so the eigenvalues are both positive, which means A defines an inner product. This is an alternate approach to showing a formula is an inner product (but is only works on finite dimensional vector spaces, so the FCC example can't be done this way). Example 7.15 in the book contains some more information here.

Example 5. An inner product on the set of m by n matrices $M_{m,n}$ is found by computing $\langle A, B \rangle = trace(B^T A)$. The examples in the text on page 247 are sufficient here. Please get practice by working with inner products that are not the usual dot product.

3 Orthogonal Complements and Projections

Recall that two vectors in an inner product space are said to be orthogonal if their inner product is zero, $\langle \vec{u}, \vec{v} \rangle = 0$. Given a single vector $\vec{u} = (u_1, \ldots, u_n)$ in $\mathbb{R}^n$, a vector $\vec{v} = (v_1, \ldots, \vec{v}_n)$ is orthogonal to $\vec{u}$ if and only if $u_1(v_1) + \cdots + u_n(v_n) = 0$. If we think of the elements of $\vec{u}$ as coefficients of a single equation and the elements of $\vec{v}$ as variables, then $\vec{v}$ is orthogonal to $\vec{u}$ if and only if $\vec{v}$ is in the null space of the 1 by n matrix $\vec{u}^T$ (where we think of $\vec{u}$ as a column matrix). So the set of all vectors orthogonal to $\vec{u}$ is just the null space of $\vec{u}^T$. This generalizes as follows: if we want to find all vectors that are orthogonal to both $\vec{u}_1$ and $\vec{u}_2$, the answer is the null space of $\begin{bmatrix} \vec{u}_1 & \vec{u}_2 \end{bmatrix}^T$. We now make a crucial defintion.

Definition 2. The orthogonal complement of a set of vectors S in an inner product space V, denoted by $S^\perp$, is the set of all vectors that are orthogonal to every vector in S. In set notation, this is written $\{\vec{v} \in V \,|\, \langle \vec{u}, \vec{v} \rangle \text{ for all } \vec{u} \in S\}$.

Theorem 1. *The orthogonal complement of a set S of vectors is a vector space. The span of S has the same orthogonal complement as S. If A is a matrix whose columns are basis vectors for the span of S, then the orthogonal complement of S is precisely the null space of A^T.*

If the set of vectors S is finite, then we can find the orthogonal complement by just finding the null space of the transpose of the matrix whose columns are the vectors (don't worry about finding a basis, because the repeated ones will disappear through row reduction). If the set of vectors is infinite (but we are in a

finite vector space), then find a basis for the span of the vectors. Then $S^{\perp}$ is precisely the null space of the transpose of the matrix whose columns are these basis vectors. Almost always we will be working with a finite set of vectors.

Example 6. Consider the set $S = \{(1, 2, -4)\}$. Let A the matrix whose only column is the vector in S. The reduced row echelon form of A^T is $\begin{bmatrix} 1 & 2 & -4 \end{bmatrix}$. A basis for the null space is $\{(-2, 1, 0), (4, 0, 1)\}$, so $S^{\perp} = \text{span}\{(-2, 1, 0), (4, 0, 1)\}$, the two dimensional plane which meets the given vector at a 90 degree angle.

Example 7. Consider the set $S = \{(1, 0, 3), (0, 2, 4)\}$. Let A the matrix whose columns are the vectors in S. The reduced row echelon form of A^T is $\begin{bmatrix} 1 & 0 & 3 \\ 0 & 1 & 2 \end{bmatrix}$. A basis for the null space is $\{(-3, -2, 1)\}$, so $S^{\perp} = \text{span}\{(-3, -2, 1)\}$, the one dimensional line which meets the plane spanned by the given vectors at a 90 degree angle.
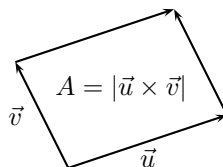
3.1   Cross Product

The cross product of two vectors $\vec{u} = (u_1, u_2, u_3)$ and $\vec{v} = (v_1, v_2, v_3)$ is a new vector

$$\vec{u} \times \vec{v} = (u_2 v_3 - u_3 v_2, -(u_1 v_3 - u_3 v_2), u_1 v_2 - u_2 v_1) = \det \begin{bmatrix} \vec{i} & \vec{j} & \vec{k} \\ u_1 & u_2 & u_3 \\ v_1 & v_2 & v_3 \end{bmatrix},$$

where $\vec{i} = (1, 0, 0), \vec{j} = (0, 1, 0), \vec{k} = (0, 0, 1)$.

1. The cross product of two vectors is a new vector which is orthogonal to both $\vec{u}$ and $\vec{v}$. This can be checked by performing the dot products $\vec{u} \cdot (\vec{u} \times \vec{v})$ and $\vec{v} \cdot (\vec{u} \times \vec{v})$. Hence, the cross product is a vector in the orthogonal complement of $\vec{u}$ and $\vec{v}$. The cross product is a special vector found in the orthogonal complement of a plane. It is studied in more depth in multivariable calculus, and used by engineers and physicist in many applications.

2. The magnitude of the cross product is $|\vec{u} \times \vec{v}| = |\vec{u}||\vec{v}| \sin \theta$, which is very similar to the dot product, but it involves a $\sin \theta$ instead of $\cos \theta$. Using this formula, the length of the cross product is the area of the parallelogram formed using the two vectors $\vec{u}$ and $\vec{v}$.

3. Since the cross product is orthogonal to both $\vec{u}$ and $\vec{v}$, it must point in one of two directions. The direction of the cross product is found using the right hand rule. If you place your right index finger on the vector $\vec{u}$, and then rotate your hand so that your middle finger is in the direction of $\vec{v}$, then the direction of $\vec{u} \times \vec{v}$ is in the direction of your thumb.

4. Note that $\vec{u} \times \vec{v} = -\vec{v} \times \vec{u}$. Because of this, we say the cross product is anti-commutative, so be careful not to switch the order on the cross product.

Example 8. A vector which is orthogonal to both $(1, -2, 3)$ and $(2, 0, -1)$ is $(1, -2, 3) \times (2, 0, -1) = \det \begin{bmatrix} \vec{i} & \vec{j} & \vec{k} \\ 1 & -2 & 3 \\ 2 & 0 & -1 \end{bmatrix} = ((-2)(-1) - (0)(3), -[(1)(-1) - (2)(3)], (1)(0) - (2)(-2)) = (2, 7, 4)$. Notice that if I reverse the order, then $(2, 0, -1) \times (1, -2, 3) = -(2, 7, 4)$ is also orthogonal to both $(1, -2, 3)$ and $(2, 0, -1)$, it just points in the opposite direction. In addition, the area of the parallelogram formed by the vectors $(1, -2, 3)$ and $(2, 0, -1)$ is $|(2, 7, 4)| = \sqrt{4 + 49 + 16} = \sqrt{69}$.

(Draft March 25, 2009)

3.2   Projections

The projection of $\vec{v}$ onto the vector $\vec{w}$, written $\text{proj}_{\vec{w}}\vec{v}$ is a vector parallel to $\vec{w}$ whose magnitude is the component of $\vec{v}$ in the direction of $\vec{w}$. If you draw both vectors with their base at the origin, and create a right triangle with $\vec{v}$ as the hypotenuse and the adjacent edge on the line containing $\vec{w}$, then $\text{proj}_{\vec{w}}\vec{v}$ is the vector which starts at the origin and ends where the right triangle ends.
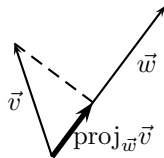
The length of the projection (using right triangle trigonometry in 2D) is $|\vec{v}|\cos\theta = |\vec{v}|\dfrac{\vec{v}\cdot\vec{w}}{|\vec{v}||\vec{w}|} = \dfrac{\vec{v}\cdot\vec{w}}{|\vec{w}|}$. A unit vector for the direction of the projection is $\dfrac{\vec{w}}{|\vec{w}|}$. Hence we have

$$\text{proj}_{\vec{w}}\vec{v} = \left(\frac{\vec{v}\cdot\vec{w}}{|\vec{w}|}\right)\frac{\vec{w}}{|\vec{w}|} = \left(\frac{\vec{v}\cdot\vec{w}}{\vec{w}\cdot\vec{w}}\right)\vec{w}.$$

In terms of inner products, the projection of $\vec{v}$ onto $\vec{w}$ is a vector $c\vec{w}$ in the direction of $\vec{w}$ which is "closest" to $\vec{v}$. In other words, I want the difference $\vec{v} - c\vec{w}$ to be as small as possible, which means that $\vec{v} - c\vec{w}$ is orthogonal to $\vec{w}$. Since it is orthogonal, this means

$$\langle \vec{v} - c\vec{w}, \vec{w}\rangle = 0 \Rightarrow \langle\vec{v},\vec{w}\rangle - c\langle\vec{w},\vec{w}\rangle = 0 \Rightarrow \langle\vec{v},\vec{w}\rangle = c\langle\vec{w},\vec{w}\rangle \Rightarrow \text{proj}_{\vec{w}}\vec{v} = c\vec{w} = \frac{\langle\vec{v},\vec{w}\rangle}{\langle\vec{w},\vec{w}\rangle}\vec{w}.$$

The quantity $c = \frac{\langle\vec{v},\vec{w}\rangle}{\langle\vec{w},\vec{w}\rangle}$ is called the Fourier coefficient of $\vec{v}$ with respect to $\vec{w}$, or the component of $\vec{v}$ along $\vec{w}$.

Example 9. The projection of $\vec{v} = (0,3,4)$ onto $\vec{w} = (2,1,-2)$ is $\text{proj}_{(2,1,-2)}(0,3,4) = \dfrac{\langle(0,3,4),(2,1,-2)\rangle}{\langle(2,1,-2),(2,1,-2)\rangle}(2,1,-2) = \dfrac{-5}{9}(2,1,-2)$. If we reverse the order, then the projection of $\vec{w} = (2,1,-2)$ onto $\vec{v} = (0,3,4)$ is $\text{proj}_{(0,3,4)}(2,1,-2) = \dfrac{\langle(2,1,-2),(0,3,4)\rangle}{\langle(0,3,4),(0,3,4)\rangle}(0,3,4) = \dfrac{-5}{25}(0,3,4)$. In Maple you can type in vectors and it will illustrate the projection for you.

3.3   Projections onto Subspaces and Linear Regression

The projection of a vector $\vec{b}$ onto a subspace W is the vector $\vec{w}$ in that space whose head is nearest to the head of $\vec{b}$. To find this vector, we seek a vector $\vec{n}$ in the orthogonal complement of W such that $\vec{b} - \vec{n} = \vec{w}$ is a vector in W. To achieve this, let A be a matrix whose columns form a basis for W. Then the solutions to $A^T\vec{n} = \vec{0}$ are precisely the vectors in the orthogonal complement of W. Because $\vec{b} - \vec{n} = \vec{w}$ is in the space W, we know that $A\vec{x} = \vec{w}$ for some vector $\vec{x}$ (as $\vec{w}$ is in the column space of A, and $\vec{x}$ gives the coordinates of $\vec{w} = \vec{b} - \vec{n}$ relative to the basis of W). Multiplying both sides by A^T gives us

$$A^TA\vec{x} = A^T\vec{w} = A^T\vec{b} - A^T\vec{n} = A^T\vec{b}.$$

In other words, the vectors $\vec{x}$ and $\vec{b}$ satisfy $A^TA\vec{x} = A^T\vec{b}$. Does this look familiar? Does it look like our linear regression formula? This is precisely where it comes from. Linear regression is just projecting a vector onto a subspace, finding the components $\vec{x}$ of the vector $\vec{w}$ in the space W that is nearest to $\vec{b}$.

Example 10. Let's find a line that is closest to passing through the three points $(0,1),(2,3),(4,6)$. Suppose for a moment that there were a line of the form $y = mx + b$ that did pass through the points. Plugging our

3 points into the equation $mx + b = y$ gives the inconsistent system of equations

$$\begin{cases} b = 1 \\ 2m + b = 3 \\ 4m + b = 6 \end{cases} \xrightarrow{\text{augmented matrix}} \begin{bmatrix} 0 & 1 & | & 1 \\ 2 & 1 & | & 3 \\ 4 & 1 & | & 6 \end{bmatrix} \xrightarrow{\text{matrix form}} \begin{bmatrix} 0 & 1 \\ 2 & 1 \\ 4 & 1 \end{bmatrix} \begin{bmatrix} m \\ b \end{bmatrix} = \begin{bmatrix} 1 \\ 3 \\ 6 \end{bmatrix}.$$

Notice that the system can be written in matrix form $A\vec{x} = \vec{b}$ where A contains a column of x values and 1's and $\vec{b}$ is a column of y values. Because the system above has no solution, we want to find the vector $\vec{w} = \vec{b} - \vec{n}$ in the column space of A (meaning $\vec{w} = A\vec{x}$)where $\vec{w}$ is the projection of $\vec{b}$ onto the 2D space spanned by the columns of A. In other words, find the shortest vector $\vec{n}$ such that $\vec{w} = \vec{b} - \vec{n}$ is a solution to $A\vec{x} = \vec{b} - \vec{n}$. The vector $\vec{n}$ must exist, so multiply both sides by A^T to obtain

$$A^T A\vec{x} = A^T(\vec{b} - \vec{n}) = A^T\vec{b} - A^T\vec{n} = A^T\vec{b},$$

where $A^T\vec{n} = \vec{0}$ because it is in the orthogonal complement of the columns of A (the shortest vector from $\vec{b}$ to the column space of $\vec{A}$ must be orthogonal to the columns of A). The resulting system has only 2 rows and a symmetric coefficient matrix $A^T A$:

$$A = \begin{bmatrix} 0 & 1 \\ 2 & 1 \\ 4 & 1 \end{bmatrix}, A^T = \begin{bmatrix} 0 & 2 & 4 \\ 1 & 1 & 1 \end{bmatrix}, A^T A = \begin{bmatrix} 20 & 6 \\ 6 & 3 \end{bmatrix}, A^T\vec{b} = \begin{bmatrix} 30 \\ 10 \end{bmatrix}, \begin{bmatrix} 20 & 6 \\ 6 & 3 \end{bmatrix} \begin{bmatrix} m \\ b \end{bmatrix} = \begin{bmatrix} 30 \\ 10 \end{bmatrix}.$$

Reduce $\begin{bmatrix} 20 & 6 & | & 30 \\ 6 & 3 & | & 10 \end{bmatrix}$ to $\begin{bmatrix} 1 & 0 & | & 5/4 \\ 0 & 1 & | & 5/6 \end{bmatrix}$, which means the solution is $y = \frac{5}{4}x + \frac{5}{6}$. Alternatively, invert A to obtain the solution

$$\vec{x} = (A^T A)^{-1} A^T\vec{b} = \begin{bmatrix} 1/8 & -1/4 \\ -1/4 & 5/6 \end{bmatrix} \begin{bmatrix} 30 \\ 10 \end{bmatrix} = \begin{bmatrix} 5/4 \\ 5/6 \end{bmatrix}.$$

4   Orthogonal and Orthonormal

The least squares regression problem introduced an idea related to finding the projection of a vector $\vec{b}$ onto a vector subspace W. This requires that we obtain a basis for the space W. Once we have picked a basis, then the projection $A\vec{x} = \vec{w}$ must satisfy $A^T A\vec{x} = A^T\vec{b}$, where $\vec{x}$ is the coordinates of $\vec{w}$ relative to our basis. Solving for $\vec{x}$ we obtain $\vec{x} = (A^T A)^{-1} A^T\vec{b}$. To find the projection $\vec{w} = A\vec{x}$, we multiply by A to obtain

$$\vec{w} = \text{proj}_W \vec{b} = A\vec{x} = A(A^T A)^{-1} A^T\vec{b}.$$

The coordinates of $\vec{w}$ relative to this basis are $\vec{x} = (A^T A)^{-1} A^T\vec{b}$, and are called the Fourier coefficients of $\vec{b}$ relative to the basis chosen for W. In this section we will show how to obtain a "nice" basis so that obtaining the projection and it's coordinates is a simple process.

The standard basis vectors $(1,0,0), (0,1,0), (0,0,1)$ are each orthogonal to the other, so we say they are an orthogonal set of vectors. These standard basis vectors are also all unit vectors. Because they are all unit vectors and orthogonal, we call them an orthonormal set of vectors. As we proceed throughout this section, we will find that when you choose as your basis an orthonormal set of vectors, computations become extremely simple (hence computer programmers prefer to work with orthonormal bases). The telecommunications industry works with an orthonormal basis because computations are extremely fast with this basis.

Example 11. The set of vectors $(2, -1, 0), (1, 2, 0), (0, 0, 1)$ forms an orthogonal set because the dot product of any two vectors is zero. Normalizing each vector gives the set $(\frac{2}{\sqrt{5}}, -\frac{1}{\sqrt{5}}, 0), (\frac{1}{\sqrt{5}}, \frac{2}{\sqrt{5}}, 0), (0, 0, 1)$, which is an orthonormal set of vectors.

In terms of projections, if we choose basis vectors $S = \{\vec{w}_1, \ldots, \vec{w}_n\}$ for W that are orthogonal, then $A^T A$ is a diagonal matrix (the diagonal entries are the square of the lengths of the basis vectors), and the coordinates of our projection relative to this basis (the Fourier coefficients) are $\vec{x} = (A^T A)^{-1} A^T\vec{b}$. Inverting

a diagonal matrix is just inverting each diagonal entry, so the coordinates $\vec{x} = (c_1, \ldots, c_n)$ are found by computing the inner product of each row with $\vec{b}$ and dividing by the length of that row, or

$$c_i = \frac{\vec{w}_i^T \vec{b}}{\vec{w}_i^T \vec{w}_i} = \frac{\vec{w}_i \cdot \vec{b}}{\vec{w}_i \cdot \vec{w}_i} = \frac{\langle \vec{w}_i, \vec{b} \rangle}{\langle \vec{w}_i, \vec{w}_i \rangle} \qquad \text{(Fourier coefficients)}.$$

If we choose basis vectors S that are orthonormal, then $A^T A$ is the identity matrix. This means $\vec{w} = AA^T \vec{b}$. The coordinates of $\vec{w}$ relative to this basis (called Fourier coefficients) are simply $\vec{x} = A^T \vec{b}$. Notice that this means the coordinates $\vec{x} = (c_1, \ldots, c_n)$ are found by computing the inner product of each row with $\vec{b}$ (no division is needed because each row has length 1), or $c_i = \vec{w}_i^T \vec{b} = \vec{w}_i \cdot \vec{b}$.

Example 12. What is the projection of $(3, -2, 5)$ onto the space spanned by the vectors $\vec{v}_1 = (2, -1, 0)$ and $\vec{v}_2 = (1, 2, 0)$? Since the vectors are orthogonal, we can find the Fourier coefficients by dotting each $\vec{v}_i$ with $\vec{b}$ and dividing by the square of the length. We obtain

$$c_1 = \frac{\langle \vec{v}_1, \vec{b} \rangle}{\langle \vec{v}_1, \vec{v}_1 \rangle} = \frac{\langle (2, -1, 0), (3, -2, 5) \rangle}{\langle (2, -1, 0), (2, -1, 0) \rangle} = \frac{8}{5}, c_2 = \frac{\langle \vec{v}_2, \vec{b} \rangle}{\langle \vec{v}_2, \vec{v}_2 \rangle} = \frac{\langle (1, 2, 0), (3, -2, 5) \rangle}{\langle (1, 2, 0), (1, 2, 0) \rangle} = \frac{-1}{5}.$$

This means the projection is $\vec{w} = c_1 \vec{v}_1 + c_2 \vec{v}_2 = \frac{8}{5}(2, -1, 0) + \frac{-1}{5}(1, 2, 0) = \frac{(16-1, -8-2, 0)}{5} = (3, -2, 0)$.

Let's consider the same set of vector, but now use an orthonormal basis to do the projection. These vectors are orthogonal, so normalizing them yields the orthonormal set of vectors $\vec{w}_1 = \frac{1}{\sqrt{5}}(2, -1, 0), \vec{w}_2 = \frac{1}{sqrt5}(1, 2, 0)$. The coefficients of the projection (the Fourier coefficients relative to the orthonormal basis) are hence $c_1 = \vec{w}_1 \cdot \vec{b} = \frac{1}{\sqrt{5}}(2, -1, 0) \cdot (3, -2, 5) = \frac{8}{\sqrt{5}}$ and $c_2 = \vec{w}_2 \cdot \vec{b} = \frac{1}{\sqrt{5}}(1, 2, 0) \cdot (3, -2, 5) = \frac{-1}{\sqrt{5}}$. This means the projection $\vec{w}$ onto the space is

$$\vec{w} = c_1 \vec{w}_1 + c_2 \vec{w}_2 = \frac{8}{\sqrt{5}} \vec{w}_1 + \frac{1-}{\sqrt{5}} \vec{w}_2 = \frac{8}{\sqrt{5}} \frac{1}{\sqrt{5}}(2, -1, 0) + \frac{-1}{\sqrt{5}} \frac{1}{\sqrt{5}}(1, 2, 0) = \frac{1}{5}(16 - 1, -8 - 2, 0) = (3, -2, 0).$$

4.1   Finding an Orthonormal Basis - Gram-Schmidt Process

Given a set of linearly independent vectors $\{\vec{v}_1, \ldots, \vec{v}_n\}$, how do I create a set of orthonormal vectors $\{\vec{w}_1, \ldots, \vec{w}_n\}$ whose span is the same as the span of the original vectors? Start with your first vector and let $\vec{w}_1 = \vec{v}_1$. Since $\vec{v}_2$ is not in the space spanned by $\vec{w}_1$, find the projection of $\vec{v}_2$ onto $\vec{w}_1$ and then let $\vec{w}_2$ be $\vec{v}_2$ minus this projection, $w_2 = v_2 - \frac{\langle \vec{v}_2, \vec{w}_1 \rangle}{\langle \vec{w}_1, \vec{w}_1 \rangle} \vec{w}_1$. This means that $\vec{w}_1$ and $\vec{w}_2$ are orthogonal, and they span the same space as $\vec{v}_1$ and $\vec{v}_2$. Since $\vec{v}_3$ is not in the space spanned by $\vec{w}_1$ and $\vec{w}_2$, find the projection of $\vec{v}_3$ onto the plane spanned by $\vec{w}_1$ and $\vec{w}_2$ and then let $\vec{w}_3$ be $\vec{v}_3$ minus this projection, $w_3 = v_3 - \frac{\langle \vec{v}_3, \vec{w}_1 \rangle}{\langle \vec{w}_1, \vec{w}_1 \rangle} \vec{w}_1 - \frac{\langle \vec{v}_3, \vec{w}_2 \rangle}{\langle \vec{w}_2, \vec{w}_2 \rangle} \vec{w}_2$. This means that $\vec{w}_1, \vec{w}_2$, and $\vec{w}_3$ are orthogonal, and span the same space as $\vec{v}_1, \vec{v}_2, \vec{v}_3$. Repeat this process, each time projecting $\vec{v}_k$ onto the subspace generated by the $k - 1$ previous orthogonal vectors. Once you are done, you will have an orthogonal set of vectors. Divide each vector by its magnitude to obtain an orthonormal set. The text has a description of this on page 255.

Example 13. Example 7.13 in the book (page 256) illustrates this nicely. Due to time constraints, I won't have my own example here.

4.2   Using Orthogonal Matrices to Diagonalize Symmetric Matrices

A matrix is called orthogonal if $A^T A = AA^T = I$, or in other words $A^T = A^{-1}$ (the transpose is the inverse). (see page 113-114 in the text). If $\vec{u}_i$ is the ith column of A, then the product $A^T A$ has in the ijth spot the product $\vec{u}_i^T \vec{u}_j$. In order for a matrix to be orthogonal, this requires that $\vec{u}_i^T \vec{u}_j = 0$ if $i \neq j$ (so columns i and j are orthogonal) and that $\vec{u}_i^T \vec{u}_j = 1$ when $i = j$ (so each vector has length 1). We have just proved the following theorem

Theorem 2. *The following three conditions are equivalent for a square matrix A:*

1. *A is orthogonal - meaning $A^T A = A A^T = I$.*

2. *The columns of A form an orthonormal set.*

3. *The rows of A form an orthonormal set.*

A key use of this theorem is that if a similarity matrix P has orthonormal columns, then its inverse is the transpose.

Example 14. Example 7.14 in your book. Part b is a rotation matrix. Every rotation in any dimension is represented by an orthogonal matrix, so computer graphic designers use these all the time.

A symmetric matrix is a matrix which satisfies $A = A^T$. The real spectral theorem states that every symmetric matrix has real eigenvalues, the geometric multiplicity of each eigenvalue equals the algebraic multiplicity, and that eigenvectors corresponding to different eigenvectors are orthogonal. Another way of saying this is that every symmetric matrix can be diagonalized by an orthogonal matrix. The converse is also true, namely every matrix which can be diagonalized by an orthogonal matrix is symmetric. Suppose A can be diagonalized by an orthogonal matrix P. Then $D = P^{-1}AP$. Since P is orthogonal, this means $D = P^T AP$, or $PDP^T = A$. The transpose of both sides gives $PD^T P^T = A^T$, and since D is diagonal we $A^T = PD^T P^T = PDP^T = A$, so $A^T = A$ and A is symmetric. Proving that every symmetric matrix is diagonalizable is much harder, and is left to a future course in linear algebra. For now, we just want to become familiar with the the follow facts:

1. Every symmetric matrix is diagonalizable over the real numbers (meaning every eigenvalue is real, and geometric and algebraic multiplicities are equal.)

2. Eigenvectors corresponding to different eigenvalues are orthogonal.

3. Every symmetric matrix can be diagonalized by an orthogonal matrix. In other words, every symmetric linear transformation has an orthonormal basis for which its matrix representation relative to that basis is diagonal.

Example 15. Page 404, example 11.9. When finding the characteristic polynomial, you can either compute the trace and determinant to find the coefficients, or you can subtract λ from the diagonals and then find the determinant. It doesn't matter which you do.

Now, do you remember the second derivative test we learned at the beginning of the semester? Every matrix involved in the second derivative test will always be symmetric. This means that with every 2nd derivative test (optimization problem), you will always be able to find the eigenvalues. In addition, eigenvectors can be chosen to always meet at right angles. Changing basis is a way of rearranging how we view a problem, and symmetric matrices have extremely nice bases. In a future linear algebra course you would learn a lot more about this fact, and use it to fully explore how the telecommunications industry and the FCC utilize linear algebra in broadcasting.

(Draft March 25, 2009)