These are brief notes for the lecture on Friday November 13, and Monday November 16, 2009: they are not complete, but they are a guide to what I want to say on those days. They are guaranteed to be incorrect.

6.1. Inner Product, Length and Orthogonality

It's very useful to be able to talk about geometric concepts such as the length of a vector, the angle between two vectors, and the projection of one vector onto another.

The Inner Product:

The key tool will turn out to be the *inner product* of two vectors in $\mathbb{R}^n$. If we have two vectors $\underline{u}$ and $\underline{v}$ in $\mathbb{R}^n$, we can regard them as $n \times 1$ matrices. Then $\underline{u}^T$ is a $1 \times n$ matrix, and the matrix product $\underline{u}^T \underline{v}$ is a 1×1 matrix, which we can regard as a scalar.

We call $\underline{u}^T \underline{v}$ the *inner product* of $\underline{u}$ and $\underline{v}$, and we write it $\underline{u} \cdot \underline{v}$. If

$$\underline{u} = \begin{pmatrix} u_1 \\ u_2 \\ \dots \\ u_n \end{pmatrix} \qquad \text{and} \qquad \underline{v} = \begin{pmatrix} v_1 \\ v_2 \\ \dots \\ v_n \end{pmatrix}$$

then

$$\underline{u} \cdot \underline{v} = [u_1 \ u_2 \ \dots \ u_n] \begin{pmatrix} v_1 \\ v_2 \\ \dots \\ v_n \end{pmatrix} = u_1v_1 + u_2v_2 + \dots u_nv_n.$$

The reason that the inner product is useful is that it satisfies the following properties

THEOREM 1. *Let* $\underline{u}, \underline{v}, \underline{w} \in \mathbb{R}^n$ *and let* c *be a scalar. Then*

(1) $\underline{u} \cdot \underline{v} = \underline{v} \cdot \underline{u}$.

(2) $(\underline{u} + \underline{v}) \cdot \underline{w} = \underline{u} \cdot \underline{w} + \underline{v} \cdot \underline{w}$.

(3) $(c\underline{u}) \cdot \underline{v} = c(\underline{u} \cdot \underline{v})$.

(4) $\underline{u} \cdot \underline{u} \geq 0$ *and* $\underline{u} \cdot \underline{u} = 0$ *if and only if* $\underline{u} = \underline{0}$.

(5) $(c_1\underline{u}_1 + \dots + c_p\underline{u}_p) \cdot \underline{w} = c_1\underline{u}_1 \cdot \underline{w} + c_p\underline{u}_p \cdot \underline{w}$.

Example: Let

$$\underline{u} = \begin{pmatrix} 1 \\ 2 \\ 3 \\ 4 \end{pmatrix} \qquad \underline{v} = \begin{pmatrix} 1 \\ -2 \\ 0 \\ -1 \end{pmatrix}$$

Compute $\underline{u} \cdot \underline{v}$

Definition (Length). *The* length *(or* norm*) of* $\underline{v}$ *is the non-negative scalar* $\|\underline{v}\|$ *defined by*

$$\|\underline{v}\| = \sqrt{\underline{v} \cdot \underline{v}} = \sqrt{v_1^2 + v_2^2 + \cdots + v_n^2} \quad \textit{and} \quad \|\underline{v}\|^2 = \underline{v} \cdot \underline{v}.$$

This definition is chosen so that the Pythagorean theorem holds (that is, in two dimensions the length c of the vector which is the hypotenuse of a right triangle with horizontal length a and vertical height b satisfies $a^2 + b^2 = c^2$.

For any scalar c, $\|c\underline{v}\| = |c| \; \|\underline{v}\|$. A vector of length 1 is called a unit vector, and if $\underline{v} \neq \underline{0}$ then

$$\tfrac{1}{\|\underline{v}\|}\underline{v}$$

is a unit vector (and is in the same direction as $\underline{v}$). This is called normalizing. We will often want to take a basis for a subspace and replace it with normalized vectors.

Example: Find a unit vector which is in the same direction as $\underline{v} = \begin{pmatrix} 1 \\ 1 \\ 2 \end{pmatrix}$.

Regarding $\underline{u}$ and $\underline{v}$ as points in $\mathbb{R}^n$, we can compute the distance between $\underline{u}$ and $\underline{v}$.

Definition. *For* $\underline{u}$ *and* $\underline{v}$ *in* $\mathbb{R}^n$*, the* distance between $\underline{u}$ and $\underline{v}$, *written as* $dist(\underline{u}, \underline{v})$ *is the length of the vector* $\underline{u} - \underline{v}$. *That is,*

$$dist(\underline{u}, \underline{v}) = \|\underline{u} - \underline{v}\|.$$

Example: Compute the distance $\text{dist}(\begin{pmatrix} 7 \\ 2 \end{pmatrix}, \begin{pmatrix} 4 \\ 3 \end{pmatrix})$.

Orthogonal Vectors

We now generalize the concept of perpendicularity: however, perpendicular is defined with respect to an underlying geometry, and often carries a sense of "vertical" in colloquial speech: since we wish to extend the idea to a much broader sense, we use the special word "orthogonal" instead. However, for our purposes, othogonal and perpendicular will mean the same thing.

Definition. *Two vectors in* $\mathbb{R}^n$ *are* orthogonal *if and only if* $\underline{u} \cdot \underline{v} = 0$.

Why should this be a good definition? If $\underline{u}$ and $\underline{v}$ are perpendicular then the Pythagorean theorem says that

$$\|\underline{u}\|^2 + \|\underline{v}\|^2 = \text{dist}(\underline{u} - \underline{v})^2$$

so that

$$\begin{aligned} \|\underline{u}\|^2 + \|\underline{v}\|^2 &= \|\underline{u} - \underline{v}\|^2 \\ &= (\underline{u} - \underline{v}) \cdot (\underline{u} - \underline{v}) \\ &= \underline{u} \cdot \underline{u} - \underline{u} \cdot \underline{v} - \underline{v} \cdot \underline{u} + \underline{v}\underline{v} \\ &= \|\underline{u}\|^2 + \|\underline{v}\|^2 - 2\underline{u} \cdot \underline{v} \end{aligned}$$

so that the Pythagorean theorem is satisfied if and only if $\underline{u}$ and $\underline{v}$ are orthogonal.

If we have two vectors $\underline{u}$ and $\underline{w}$, then we can express $\underline{w}$ as a multiple of $\underline{u}$ plus a vector orthogonal to $\underline{u}$. To do this, we wish to find $\underline{v} \in \mathbb{R}^n$ and $c \in \mathbb{R}$ so that

$$\underline{w} = c\underline{u} + \underline{v} \qquad \text{and} \qquad \underline{u} \cdot \underline{v} = 0.$$

So, eliminating $\underline{v}$ in these, we obtain

$$\underline{u} \cdot (\underline{w} - c\underline{u}) = 0$$

and this implies that

$$c = \frac{\underline{w} \cdot \underline{u}}{\underline{u} \cdot \underline{u}} = \frac{\underline{w} \cdot \underline{u}}{\|\underline{u}\|^2}.$$

Hence

$$\underline{v} = \underline{w} - \frac{\underline{w} \cdot \underline{u}}{\|\underline{u}\|^2}\underline{u}.$$

We call $c\underline{u}$ the projection of $\underline{v}$ onto $\underline{u}$, and $\underline{v}$ the component of $\underline{v}$ orthogonal to $\underline{u}$.

In two or three dimensions, the projection of $\underline{v}$ onto $\underline{u}$ has length $\|\underline{v}\| \cos\theta$, where θ is the angle between the vectors. Hence

$$\|\underline{v}\| \cos\theta = c\|\underline{u}\|$$

so that

$$\underline{u} \cdot \underline{v} = \|\underline{u}\|\|\underline{v}\| \cos\theta.$$

In higher dimensions than three we use this to *define* the angle between two vectors.

6.2. Orthogonal Sets

DEFINITION. *A set* $\{\underline{u}_1, \underline{u}_2, \ldots, \underline{u}_k\}$ *of non-zero vectors is said to be orthogonal if for every* $i \neq j$,

$$\underline{u}_i \cdot \underline{u}_j = 0,$$

that is, if every pair of vectors is orthogonal.

Example: Show that the following set of vectors is orthogonal:

$$\left\{ \begin{pmatrix} 3 \\ 1 \\ 1 \end{pmatrix}, \begin{pmatrix} -1 \\ 2 \\ 1 \end{pmatrix}, \begin{pmatrix} 1 \\ 4 \\ -7 \end{pmatrix} \right\}.$$

THEOREM 4. *If* $\{\underline{u}_1, \underline{u}_2, \ldots, \underline{u}_k\}$ *is an orthogonal set of vectors, then it is linearly independent. Hence it is a basis for the space that it spans.*

Proof: proof strategy: suppose that we have a non-trivial linear combination giving zero. By taking the inner product with $\underline{u}_j$, show that the coefficient of $\underline{u}_j$ must be zero. Hence the set is linearly independent.

Definition. *An orthogonal basis for a subspace* $W < \mathbb{R}^n$ *is a basis for* W *which is an orthogonal set.*

Theorem 5. *Let* $W < \mathbb{R}^n$*, and let* $S = \{\underline{u}_1, \underline{u}_2, \dots, \underline{u}_k\}$ *be an orthogonal basis for* W*. Then if the vector* $\underline{y}$ *in* W *is given in terms of the basis* S *by*

$$\underline{y} = c_1\underline{u}_1 + c_2\underline{u}_2 + \cdots + c_k\underline{u}_k$$

then

$$c_j = \frac{\underline{y} \cdot \underline{u}_j}{\underline{u}_j \cdot \underline{u}_j}$$

Proof: Proof strategy: we know that every vector in W has a unique representation as a linear combination of vectors in S, since S is a basis for W. If we take the inner product of $\underline{y}$ with $\underline{u}_j$, then since W is orthogonal, all the inner products vanish except for the terms in the theorem.

Example: We saw above that

$$S = \left\{ \begin{pmatrix} 3 \\ 1 \\ 1 \end{pmatrix}, \begin{pmatrix} -1 \\ 2 \\ 1 \end{pmatrix}, \begin{pmatrix} 1 \\ 4 \\ -7 \end{pmatrix} \right\}$$

is an orthogonal set in $\mathbb{R}^3$. Since it is linearly independent and has three vectors in it, it must be a basis for $\mathbb{R}^3$. Express the vector

$$\begin{pmatrix} 2 \\ 4 \\ 6 \end{pmatrix}$$

in terms of the vectors in S.

Orthogonal Projections

Earlier we saw how to find the component of $\underline{v}$ in the direction of $\underline{u}$ and the component orthogonal to $\underline{u}$. We revisit this idea to introduce some notation, and to extend it to projecting onto a subspace.

Given $\underline{y} \in \mathbb{R}^n$ and $\underline{u} \in \mathbb{R}^n$, find $\underline{\hat{y}}, \underline{z} \in \mathbb{R}^n$ so that

(1) $\underline{y} = \underline{\hat{y}} + \underline{z}$.

(2) $\underline{\hat{y}} = \alpha \underline{u}$ ($\alpha \in \mathbb{R}$).

(3) $\underline{u} \cdot \underline{z} = 0$.

As before, we see that

$$\underline{y} \cdot \underline{u} = \underline{\hat{y}} \cdot \underline{u} = \alpha \|\underline{u}\|^2$$

so

$$\alpha = \frac{\underline{y} \cdot \underline{u}}{\|\underline{u}\|^2}$$

and

$$\underline{z} = \underline{y} - \frac{\underline{y} \cdot \underline{u}}{\|\underline{u}\|^2} \underline{u}$$

Notation:

$$\underline{\hat{y}} = \mathrm{Proj}_L(\underline{y}) = \frac{\underline{y} \cdot \underline{u}}{\|\underline{u}\|^2} \underline{u}$$

where $L = \mathrm{Span}(\underline{u})$.

Terminology: $\underline{y} = \underline{\hat{y}} + \underline{z}$:
$\underline{\hat{y}}$ is the orthogonal projection of $\underline{y}$ onto L
$\underline{z}$ is the component of $\underline{y}$ orthogonal to L.

Example: Let $\underline{y} = \begin{pmatrix} 7 \\ 6 \end{pmatrix}$, $\underline{u} = \begin{pmatrix} 4 \\ 2 \end{pmatrix}$ and $L = \mathrm{Span}(\underline{u})$. Compute the orthogonal projection of $\underline{y}$ onto L and the component of $\underline{y}$ orthogonal to L. Plot $\underline{y}, \underline{u}, \underline{\hat{y}}$ and $\underline{z}$. Compute the distance from $\underline{y}$ to L. (Note: the subspace L here is the line through $\underline{0}$ and $\underline{u}$.)

Geometric Interpretation of Theorem 5: Let $\{\underline{u}_1, \underline{u}_2\}$ be an orthogonal basis for $\mathbb{R}^2$. Put

$$\underline{\hat{y}}_1 = \frac{\underline{y} \cdot \underline{u}_1}{\|\underline{u}_1\|^2}\underline{u}_1 = \text{Proj}_{\underline{u}_1}(\underline{y})$$
$$\underline{\hat{y}}_2 = \frac{\underline{y} \cdot \underline{u}_2}{\|\underline{u}_2\|^2}\underline{u}_2 = \text{Proj}_{\underline{u}_2}(\underline{y})$$

Then

$$\underline{y} = \underline{\hat{y}}_1 + \underline{\hat{y}}_2.$$

Orthonormal Sets

DEFINITION. *A set* $\underline{u}_1, \underline{u}_2, \ldots, \underline{u}_p$ *is called* orthonormal *if it is orthogonal and* $\|\underline{u}_i\| = 1$ *for* $1 \leq i \leq p$. *In this case, if* $W = Span(\underline{u}_1, \ldots, \underline{u}_p)$, *then the set is called an* orthonormal basis *for* W.

Example: The set $\underline{e}_1, \underline{e}_2, \ldots, \underline{e}_n$ is an orthonormal basis for $\mathbb{R}^n$.
Example: Show that the set $\underline{v}_1, \underline{v}_2, \underline{v}_3$ is an orthornormal basis for $\mathbb{R}^3$, where

$$\underline{v}_1 = \begin{pmatrix} 3/\sqrt{11} \\ 1/\sqrt{11} \\ 1/\sqrt{11} \end{pmatrix}, \qquad \underline{v}_2 = \begin{pmatrix} -1/\sqrt{6} \\ 2/\sqrt{6} \\ 1/\sqrt{6} \end{pmatrix}, \qquad \underline{v}_3 = \begin{pmatrix} 1/\sqrt{66} \\ 4/\sqrt{66} \\ -7/\sqrt{66} \end{pmatrix}.$$

Theorem 6. *An $m \times n$ matrix U has orthonormal columns if and only if $U^T U = I$.*

Proof: proof strategy: interpret the entries of $U^T U$ in terms of inner products of the columns of U.

Theorem 7. *Let U be an $m \times n$ matrix with orthonormal columns, and $\underline{x}, \underline{y} \in \mathbb{R}^n$. Then*

(1) $\|U\underline{x}\| = \|\underline{x}\|$.

(2) $(U\underline{x}) \cdot (U\underline{y}) = \underline{x} \cdot \underline{y}$.

(3) $(U\underline{x}) \cdot (U\underline{y}) = 0$ *if and only if* $\underline{x} \cdot \underline{y}$.

Note: (a) U preserves length
(b) U preserves orthonormality.
Proof: proof strategy: write $(U\underline{x}) \cdot (U\underline{y})$ as

$$(U\underline{x}) \cdot (U\underline{y}) = (U\underline{x})^T (U\underline{y}) = (\underline{x}^T U^T)(U\underline{y}) = \underline{x}^T (U^T U)\underline{y}.$$

Example: Let

$$U = \begin{pmatrix} 1/\sqrt{2} & 2/3 \\ 1/\sqrt{2} & -2/3 \\ 0 & 1/3 \end{pmatrix} \qquad \text{and} \qquad \underline{x} = \begin{pmatrix} \sqrt{2} \\ 3 \end{pmatrix}.$$

Note that U has orthonormal columns and that $\|U\underline{x}\| = \|\underline{x}\|$.

In the case where U is a an $n \times n$ matrix with orthonormal columns, we see that $U^T U = I$ so $U^T = U^{-1}$, so $UU^T = I$ as well: that is, U has orthonormal rows too!