MA 511, Session 14

Orthogonality

Definition: The (standard) inner product (also called dot product) in $\mathbb{R}^n$ is the function $\mathbb{R}^n \times \mathbb{R}^n \longrightarrow \mathbb{R}$ given by

$$(u, v) \to u^T v = \sum_{j=1}^{n} u_j v_j .$$

The norm (or length) of a vector is

$$\|v\| = \sqrt{v^T v} = \sqrt{\sum_{j=1}^{n} v_j^2} = \sqrt{v_1^2 + \cdots + v_n^2}.$$

Definition: Two vectors $u, v \in \mathbb{R}^n$ are orthogonal if $u^T v = 0$.

Theorem: Let $v_1, \ldots, v_k \in \mathbb{R}^n$ be pairwise orthogonal (i.e. $v_i^T v_j = 0$ for $1 \leq i, j \leq k$) and assume none of them is the zero vector (which means that their lengths are positive, $\|v_j\| > 0$). Then, $v_1, \ldots, v_k$ are linearly independent.

<u>Proof</u>: Suppose $c_1 v_1 + \cdots + c_k v_k = 0$. Take the inner product of each side with v_1 to obtain

$$v_1^T (c_1 v_1 + \cdots + c_k v_k) = v_1^T 0,$$

that is

$$c_1 v_1^T v_1 + \cdots + c_k v_1^T v_k = 0.$$

Using the pairwise orthogonality, the left-hand side actually has only one term:

$$c_1 v_1^T v_1 = c_1 \|v_1\| = 0,$$

and, since $\|v_1\| \neq 0$, we conclude $c_1 = 0$. Now we take the inner product of $c_1 v_1 + \cdots + c_k v_k = 0$ with v_j $(2 \leq j \leq k)$ and, using the same argument as for v_1 we conclude that $c_1 = c_2 = \cdots = c_k = 0$, which means that $v_1, \ldots, v_k$ are linearly independent.

The Triangle Inequality: for any vectors x, y we have

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

In $\mathbb{R}^n$ this is just the statement that for any triangle the length of one of its sides is less than the sum of the lengths of the other two sides.

Remark: (Pythagorean Theorem) If x and y are orthogonal, then $\|x + y\|^2 = \|x\|^2 + \|y\|^2$.

Proof: Note that

$$\begin{aligned}\|x+y\|^2 = (x+y)^T(x+y) &= x^T x + x^T y + y^T x + y^T y \\ &= x^T x + y^T y = \|x\|^2 + \|y\|^2,\end{aligned}$$

since $x^T y = y^T x = 0$.

Lemma: For any vectors x and y, we have

$$\|x - y\| \geq \Big| \|x\| - \|y\| \Big|.$$

<u>Proof</u>: We need to show that $\|x - y\| \geq \|x\| - \|y\|$ and $\|x - y\| \geq \|y\| - \|x\|$. We'll just prove the first of these inequalities, since the other one is shown exactly the same way. Using the triangle inequality,

$$\|x\| = \|y + (x - y)\| \leq \|y\| + \|x - y\|$$

which gives

$$\|x - y\| \geq \|x\| - \|y\|.$$

Remark: Note that, by definition, for any $c \in \mathbb{R}$,

$$\|cx\| = |c| \, \|x\|,$$

so that

$$\|x - y\| = \|y - x\|.$$

Let V and W be subspaces of $\mathbb{R}^n$.

Definition: V and W are <u>orthogonal subspaces</u> if $v^T w = 0$ for all $v \in V$ and all $w \in W$.

Notation: $V \perp W$.

Example: Let V be the xy-plane in $\mathbb{R}^3$ and W be the yz-plane. Any nonzero vector on the y-axis belongs to both V and W and is not orthogonal to itself (e.g. $v = (0, 1, 0) \in V$ and $w = (0, 1, 0) \in W$, but $v^T w = 1 \neq 0$). Hence, V and W are NOT orthogonal subspaces, even though geometrically we would say they are orthogonal planes (because, actually, the orthogonality of planes in $\mathbb{R}^3$ is considered by the orthogonality of their perpendicular directions!).

Example: Let V be the xy-plane in $\mathbb{R}^3$ and W be the z-axis. Then, V and W are orthogonal subspaces.

Note that

$$V = \left\{ s \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} + t \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix} \right\} \text{ and } W = \left\{ u \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} \right\}.$$

We take $v = (s, t, 0)^T \in V$ and $w = (0, 0, u)^T \in W$ and readily see that $v^T w = s0 + t0 + 0u = 0$. Thus, they are orthogonal.

Theorem: Let A be a $m \times n$ matrix. Then,

$$\mathcal{C}(A^T) \text{ is orthogonal to } \mathcal{N}(A), \quad \text{and}$$
$$\mathcal{C}(A) \text{ is orthogonal to } \mathcal{N}(A^T).$$

Proof: It is sufficient to prove the first claim, since the second one is the first applied to A^T instead of A. So, we need to show that every vector in the null space of A is orthogonal to every vector in the row space of A. Let $x \in \mathcal{N}(A)$. Then, $Ax = 0$, which means that x is orthogonal to each row of A. It follows that x is orthogonal to all linear combinations of those rows, and therefore, to every vector in the row space of A.

Definition: If V is a subspace of $\mathbb{R}^n$, then

$$V^\perp = \left\{ w \in \mathbb{R}^n : w^T v = 0 \ \text{ for all } \ v \in V \right\}$$

is called the orthogonal complement of V.

Remark: $V \cap V^\perp = \{0\}$. In fact, $v \in V \cap V^\perp$ implies $v^T v = \|v\|^2 = 0$ which means $v = 0$.

Theorem: Let $v_1, \ldots, v_k$ and $w_1, \ldots, w_l$ be bases for V and W, respectively, where V and W are subspaces of $\mathbb{R}^n$. Then, $V \perp W$ if, and only if $v_i^T w_j = 0$ for $1 \le i \le k$ and $1 \le j \le l$.

Proof: The necessity (only if) is an immediate consequence of the definition. The sufficiency (if) follows by taking linear combinations: Given any $v \in V$ and any $w \in W$, there exist scalars $c_1, \ldots, c_k$ and $d_1, \ldots, d_l$ such that $v = c_1 v_1 + \cdots + c_k v_k$ and $w = d_1 w_1 + \cdots + d_l w_l$. Then,

$$\begin{aligned} v^T w &= (c_1 v_1 + \cdots + c_k v_k)^T (d_1 w_1 + \cdots + d_l w_l) \\ &= \sum_{i=1}^{k} \sum_{j=1}^{l} c_i d_j (v_i^T w_j) = 0. \end{aligned}$$

Remark: This result says that orthogonality of subspaces needs to be checked only for their bases.

Example: Using $\dim \mathcal{C}(A^T) + \dim \mathcal{N}(A) = n$ for any $m \times n$ matrix A, we see from the orthogonality of those two subspaces of $\mathbb{R}^n$ (and the analogous one for the column space and left-null space in $\mathbb{R}^m$) that, in fact,

$$\mathcal{C}(A^T) = \mathcal{N}(A)^{\perp}, \quad \text{and} \quad \mathcal{C}(A) = \mathcal{N}(A^T)^{\perp}.$$

As a consequence, since we know that $Ax = b$ is solvable if, and only if $b \in \mathcal{C}(A)$, we may also say $Ax = b$ is solvable if, and only if b is orthogonal to everything that is orthogonal to the columns of A: $Ax = b$ is solvable if, and only if

$$b^T y = 0 \quad \text{whenever} \quad A^T y = 0.$$

Finally, consider a $m \times n$ matrix A. We have a decomposition of $\mathbb{R}^n$ into two fundamental orthogonal subspaces, $\mathcal{C}(A^T)$ and $\mathcal{N}(A)$, and also a decomposition of $\mathbb{R}^m$ into two fundamental orthogonal subspaces, $\mathcal{C}(A)$ and $\mathcal{N}(A^T)$.

Remark: Every $v \in \mathbb{R}^n$ can be (uniquely) written as $v = v_r + v_n$ with $v_r \in \mathcal{C}(A^T)$ and $v_n \in \mathcal{N}(A)$. This fact is denoted by writing $\mathbb{R}^n = \mathcal{C}(A^T) \oplus \mathcal{N}(A)$.

Similarly, $\mathbb{R}^m = \mathcal{C}(A) \oplus \mathcal{N}(A^T)$.

Now $A : \mathbb{R}^n \longrightarrow \mathbb{R}^m$ via $v \to Av$, and we see that $Av = A(v_r + v_n) = Av_r$ since $Av_n = 0$. So, A really maps its row space into its column space, and this mapping is invertible: every $w \in \mathcal{C}(A)$ can be uniquely written as $w = Av$ with $v \in \mathcal{C}(A^T)$. Indeed, let $w = s_1 c_1 + \cdots + s_n c_n$, where $s_1, \ldots, s_n$ are scalars and $c_1, \ldots, c_n$ are the columns of A. If we let $s = (s_1, \ldots, s_n)^T$, it follows that $As = w$. Moreover, suppose there are two such representations:

$w = Av_r = Av_r'$. It follows that $A(v_r - v_r') = 0$, which says that $v_r - v_r' \in \mathcal{N}(A)$. As $v_r - v_r' \in \mathcal{C}(A^T)$ too, it follows that $v_r - v_r' = 0$, i.e. $v_r = v_r'$.