Lecture 5-1

By now (as promised) we have seen and used dot products quite a bit in linear algebra. So far, they have been used primarily as a tool to prove results about matrices; we now study them in their own right. Let S be a subspace of $\mathbb{R}^n$ and let $S^\perp$, the *orthogonal complement* of S, consist of all $v \in \mathbb{R}^n$ with $v \cdot w = 0$ for all $w \in S$. We have already seen that $S \cap S^\perp = 0$, since any $v \in S \cap S^\perp$ satisfies $v \cdot v = 0$. (Note that this does not hold if $\mathbb{R}$ is replaced by $\mathbb{C}$, wince we can have $v \cdot v = 0$ for nonzero $v \in \mathbb{C}^n$. One should therefore replace the dot product in this case by the Hermitian inner product (v, w).) Fixing a basis $b_1, \ldots, b_m$ of S, we have $w \in S^\perp$ if and only if $b_i \cdot w = 0$ for all i, whence $S^\perp$ may be regarded as the solution space to a homogeneous linear system with m independent equations, which must have $n - m$ free variables. Hence $\dim S^\perp = n - m = n - \dim S$. Combining our basis $b_1, \ldots, b_m$ of S with a basis $c_1, \ldots, c_{n-m}$ of $S^\perp$, we get an independent set, since if any combination c of the c_i equals a combination of the b_i, then $c \in S \cap S^\perp = 0$. This set must therefore be a basis. The upshot of this discussion is that *any vector in* $\mathbb{R}^n$ *may be uniquely written as the sum* $s + t$ *of some* $s \in S, t \in S^\perp$. We call s the *orthogonal projection of* v *to* S. Then s the unique vector in S closest to v, for if $s' \in S$ then the square length $(v - s') \cdot (v - s') = (s - s' + t) \cdot (s - s' + t) = ||s - s'||^2 + ||t||^2$, since $s - s'$ is orthogonal to t.

Suppose now that we are given an *inconsistent* system $AX = B$ with A an $m \times n$ matrix. Then we know that B fails to lie in the column space C of A. By the preceding paragraph, there is a unique $B' = AX$ in S closest to B which is determined by the condition that $B - B'$ be orthogonal to C, or equivalently that $A^T(B - B') = 0$, since the rows of A^T are the columns of A. Moreover, letting R be the row space of A, we know that R and C have the same dimension and that multiplication by A sends R to C (since R lives in $\mathbb{R}^n$ and A sends all of $\mathbb{R}^n$ to C). Now if any $X \in R$ satisfies $AX = 0$, then X is orthogonal to every row of A, whence $X \in R \cap R^\perp = 0$. Consequently multiplication by A defines a 1-1 linear map from R into C, which must therefore be onto.

A remarkable consequence of the above analysis is that there is a simple uniform way to replace any inconsistent system by a consistent one. Given an inconsistent system $AX = B$, the *normal equations* for this system are the ones forming the system $A^TAX = A^TB$. This last system is indeed always consistent, since it is equivalent to $A^T(AX - B) = 0$, which is equivalent to saying that $AX - B$ is orthogonal to the column space of A. Thus there is always a unique $B' = AX$ satisfying this system, namely the orthogonal projection of B to the column space of A. This projection is closer than any other AX to B If A has full rank, then A^TA is invertible and the normal equations have a unique solution for X. In general, there can be many X's with the same $B' = AX$, but there will always be a unique *shortest shortest* such X, which lies in the row space of A. We call any solution X to the normal equations a *least-squares solution* to $AX = B$. For example, a common reason in practice that a system $AX = B$ is inconsistent is that it has (many) more equations than unknowns; we call such a system *overdetermined*. The least squares solution is the most commonly used approximate solution in practice.

For example, given the system $AX = \begin{pmatrix} 1 & 0 \\ 1 & 1 \\ 2 & 1 \end{pmatrix} X = \begin{pmatrix} 2 \\ 3 \\ 4 \end{pmatrix}$, we form the augmented matrix $\begin{pmatrix} 1 & 0 & 2 \\ 1 & 1 & 3 \\ 2 & 1 & 4 \end{pmatrix}$ and row-reduce it to echelon form, obtaining $\begin{pmatrix} 1 & 0 & 2 \\ 0 & 1 & 1 \\ 0 & 0 & -1 \end{pmatrix}$, so that the system is inconsistent. Multiplying both sides of its by the transpose $\begin{pmatrix} 1 & 1 & 2 \\ 0 & 1 & 1 \end{pmatrix}$, we get the normal equations $\begin{pmatrix} 6 & 3 \\ 3 & 2 \end{pmatrix} X = \begin{pmatrix} 13 \\ 7 \end{pmatrix}$, whose unique solution is $\begin{pmatrix} 5/3 \\ 1 \end{pmatrix}$. Plugging these values for the variables back into the original system, we find that $AX = \begin{pmatrix} 5/3 \\ 8/3 \\ 13/3 \end{pmatrix}$, which comes fairly close to the vector $\begin{pmatrix} 2 \\ 3 \\ 4 \end{pmatrix}$ we really wanted.

A common application is the following one. Suppose we are interested in how one measurable quantity y depends on another one x. We perform $n+1$ experiments, setting x equal to $n+1$ distinct values $x_1, \ldots, x_{n+1}$ and measure y each time, obtaining $n+1$ values $y_1, \ldots, y_{n+1}$ (not necessarily distinct). Then a basic theorem asserts that *there is a unique polynomial p of degree at most n such that $p(x_i) = y_i$ for all i.* This result may not be of much use if say $n+1 = 1001$, for polynomials of degree 1000 are quite difficult to understand or even write down. One often asks instead for a polynomial of low degree, say 1 or 2, which best fits the data points (x_i, y_i) in the sense of least squares. We will learn how to find such a polynomial next time by solving the normal equations of an inconsistent system.

We close by mentioning that one can play the same game with subspaces S of $\mathbb{C}^n$ rather than $\mathbb{R}^n$, using the Hermitian inner product rather than the dot product, defining for any such S its orthogonal complement $S^\perp$ to consist of all $w \in \mathbb{C}^n$ with $(v, w) = 0$ for all $v \in S$; then any $u \in \mathbb{C}^n$ is equal to the sum $v + w$ of a unique $v \in S, w \in S^\perp$. In practice, however, this setting occurs much more rarely than that of $\mathbb{R}^n$.