Lecture 5-4

We wrap up our treatment of linear algebra by solving the polynomial problem posed last time and proving one more general result about $n \times n$ matrices. We stated last time that given any $n+1$ distinct numbers $x_1, \ldots, x_{n+1}$ and arbitrary numbers $y_1, \ldots, y_{n+1}$, there is a unique polynomial p of degree at most n such that $p(x_i) = y_i$ for all i. To prove this, write the polynomial p as $\sum_{j=0}^{n} a_j x^j$. Then the requirement that $p(x_i) = y_i$ reduces to the linear system $MX = B$, where M is the *Vandermonde* matrix $\begin{pmatrix} 1 & x_1 & \cdot & \cdot & \cdot & x_1^n \\ \cdot & \cdot & & & & \cdot \\ \cdot & \cdot & & & & \cdot \\ \cdot & \cdot & & & & \cdot \\ 1 & x_{n+1} & \cdot & \cdot & \cdot & x_{n+1}^n \end{pmatrix}, X = \begin{pmatrix} a_0 \\ \cdot \\ \cdot \\ \cdot \\ a_n \end{pmatrix}, B = \begin{pmatrix} y_1 \\ \cdot \\ \cdot \\ \cdot \\ y_{n+1} \end{pmatrix}$; note that the x_i and y_i are known constants here while the a_i are the unknown coefficients in p that we are trying to determine. In HW last week you showed that the determinant of this matrix is $\prod(x_j - x_i) \neq 0$, where the product takes place over all indices i, j with $1 \leq i < j \leq n+1$, since the x_i are distinct. Hence this system indeed has a unique solution, as desired. Now, as we mentioned last time, if n is large, it is much more common in practice to look for a low-degree polynomial q, say of degree at most 2, that comes as close as possible to satisfying $q(x_i) = y_i$ for all i. The same argument that showed that our matrix M is nonsingular also shows that if M is replaced by the analogous matrix with $m > n+1$ numbers x_i (each paired with a y_i) but still only $n+1$ columns, with the x_i raised to the n-th power in the last column, then M has full rank. Hence the (typically) inconsistent system $MX = B$ is such that if it is replaced by the normal equations $M^T MX = M^T B$ then these normal equations have a unique solution X. Hence there is a unique least-squares best fit q to the data points (x_i, y_i) for $1 \leq i \leq m$, usually not passing through any of these points, but coming fairly close to all of them. The uniqueness of q in this situation is very convenient in applications. (Other measures of goodness of fit, different from the least-squares one, are also sometimes used in practice, but much less often.)

Returning to the theory one last time we now let A be an $n \times n$ matrix with characteristic polynomial $p(\lambda) = \det(A - \lambda I)$; as previously observed, p is a polynomial of degree n. Since all powers $A^0 = I, A, A^2, \ldots$ make sense and are again $n \times n$ matrices, it makes sense to evaluate the polynomial p at the matrix A rather than a number. The famous *Cayley-Hamilton Theorem* asserts that $p(A) = 0$ (the 0 matrix). To prove this we let B_λ be the transpose of the cofactor matrix of $A - \lambda I$. Regard both $A - \lambda I$ and B_λ as matrices with entries in $\mathbb{R}[\lambda]$, the ring of polynomials in one variable λ. We then have $B_\lambda(A - \lambda I) = (A - \lambda I)B_\lambda = p(\lambda)I$, by the same formal argument that showed that if any $n \times n$ matrix M is multiplied on the left by the transpose N of its cofactor matrix than the product $NM = (\det M)I$. Writing B_λ as a polynomial $\sum_i M_i \lambda^i$ with matrix coefficients (where the M_i have constant entries, independent of λ), we see by induction on i from this calculation that the matrices M_i all commute with A. Now given any polynomial $q(\lambda) = \sum N_i \lambda^i$ with constant matrix coefficients N_i commuting with A, we can substitute A for λ. In particular, doing this with $A - \lambda I$ gives the zero matrix, while doing with with $p(\lambda)I$ gives $p(A)$. Hence $p(A) = 0$, as claimed.

In particular, any $n \times n$ matrix satisfies a polynomial of degree n and leading coefficient $(-1)^n$. The same argument works for matrices over any field K in place of $\mathbb{R}$, and in fact for matrices over any commutative ring; we never needed to divide by any entry of a matrix. Thus for example an $n \times n$ matrix over the ring $\mathbb{Z}$ of integers satisfies a polynomial of degree n and leading coefficient $(-1)^n$ with *integer* coefficients. Multiplying this polynomial by -1 if necessary, we may assume that its leading coefficient is 1. Almost all of the results about matrices that we have proved this term, though stated only for matrices with real entries, in fact apply to matrices with entries in any field. In abstract algebra, one generalizes many of these results to matrices with entries in a commutative ring; the Cayley-Hamilton Theorem is a prime example.

We also mention that Cramer's Rule (giving the unique solution to a linear system $AX = B$ with A square and invertible, such that the ith coordinate x_i of the unique solution X is a ratio of determinants), although quite cumbersome to apply in practice to solve linear systems, is extremely important theoretically; it is used to state and prove the general Implicit Function Theorem for m functional equations $f_1(x_1, \ldots, x_{n+m}) = \ldots = f_m(x_1, \ldots, x_{n+m}) = 0$ in $m+n$ variables x_i, using these equations to solve for m of the x_i differentiably in terms of the others (if the f_i have continuous partials). It is very handy to have explicit uniform formulas for the partial derivatives of the resulting m of the x_i with respect to the others and that is just what Cramer's Rule provides, expressing these partial derivatives as ratios of determinants.

We wrap up the course with multivariable integral calculus, following Chapter 17 of Salas-Hille. Linear algebra will make a final grand appearance later, in the context of the change of variable formula for multiple integrals.