Lecture 4-23

We continue with reviewing material covered so far, proceeding now to linear algebra. We began with systems of linear equations; we have seen how, given such a system $MX = B$ we can perform row operations on the augmented matrix M' obtained from M by adding B as a new column to its right so as to bring M to echelon form, so that the leftmost nonzero entry of every nonzero row of M has zeroes below it, the position of this leftmost entry moves strictly to the right as you go down the rows, and all zero rows of M are grouped together at the bottom. The resulting system is then consistent (has a solution) if and only if no zero row in the echelon form of M has a nonzero rightmost entry. If this is the case, then we can write down the entire solution set by identifying the free variables of the system (corresponding to columns of M not having pivots in its echelon form), assigning arbitrary values to these variables, and then solving for the other variables in terms of the free ones.

Turning now to the abstract linear algebra behind the algorithm for solving systems, we say that a collection $v_1, \ldots, v_m$ of vectors in a vector space V is (linearly) independent if the only combination $\sum a_i v_i$ equal to the 0 vector has $a_i = 0$ for all i. The span of the same set of vectors in V consists by definition of all such combinations $\sum a_i v_i$ and is a subspace of V. A system $MX = B$ has a solution if and only if B lies in the column span of M (the subspace spanned by the columns); in turn the columns of M are linearly independent if and only if the only solution to the homogeneous system $MX = 0$ is the 0 solution. Thus, in particular, any set of more than n vectors in a vector space of dimension n (e.g. $\mathbb{R}^n$) is necessarily dependent, since the matrix M formed by the coordinates of its columns (with respect to some basis) cannot have a pivot in every column in its echelon form, so that the system $MX = 0$ must have at least one free variable. By the same token, no set of fewer than n vectors can span $\mathbb{R}^n$, or any vector space of dimension n. Any basis of $\mathbb{R}^n$ has exactly n vectors, and more generally any two bases of the same vector space have the same number of elements, called the dimension of that space. Given a vector space of dimension n, any set of n vectors spans the space if and only if it is linearly independent, so that the two requirements in the definition of basis (it must span the space *and* be independent) are equivalent for any set of n vectors in a space of dimension n.

There are three important subspaces of $\mathbb{R}^p$ (for various p) attached to any (possibly rectangular) $m \times n$ matrix M, namely its *row space*, spanned by its rows and living in $\mathbb{R}^n$; its *column space*, spanned by its columns and living in $\mathbb{R}^m$; and its *kernel* or *nullspace*, consisting of all column vectors $V \in \mathbb{R}^n$ with $MV = 0$. The row and columns spaces always have the same dimension, called the *rank* of M; the dimension of the kernel of M equals the number of its columns minus its rank. (A fourth subspace, called the *left nullspace* and consisting of all *row* vectors $V \in \mathbb{R}^m$ with $VM = 0$, is less important; its dimension equals the number of rows of M minus its rank.)

Given two finite-dimensional vector spaces V, W with fixed respective bases B, B', any linear transformation f from V to W (such that $f(v_1 + v_2) = f(v_21) + f(v_2), f(rv) = rf(v)$ for all $v, v_1, v_2 \in V, r \in \mathbb{R}$) has a unique matrix A, whose ith column consists of the coefficients of $f(b_i)$, the image of the i-th vector of B, when written as a combination of the vectors in B'; we can then compute $f(v)$ for any $v \in V$ by writing v as a linear combination of vectors in B, assembling the coefficients of the vectors of B into a column vector, multiplying A by thie column vector, and interpreting the resulting column vector as a combination of vectors in B', taking the coordinates of this vector as the coefficients in the combination. If $V = \mathbb{R}^n, W = \mathbb{R}^m$, then we can say more snappliy that *any linear transformation f from V to W is given by left multiplication of column vectors of length n to a unique $m \times n$ matrix, producing thereby a column vector of length m.* The column space of the matrix then coincides with the range of f, while its nullspace is the same as the kernel of f, defined to be the space of all $v \in \mathbb{R}^n$ with $f(v) = \vec{0}$.

Given any real *symmetric* matrix M (equalling its own transpose), we can perform row operations on it as usual to bring it to echelon form, but follow each by the corresponding column operation, so as to preserve the symmetric nature of M throughout. In the end we produce a diagonal matrix D. The original matrix M is *positive definite* in the sense that $\vec{V}^T MV > 0$ for all nonzero column vectors V of the appropriate length if and only if all diagonal entries of D are positive; similarly M is negative definite if and only if all diagonal entries of D are negative.

Finally, given any matrices A, B for which the products AB, BA are both defined (and thus both square), we have tr $AB =$ tr BA, where the *trace* tr C of any square matrix C is defined to be the sum of the entries along its main diagonal. More generally, if a matrix product AB is defined, so is the reverse product $B^T A^T$ of the transposes B^T and A^T of B and A and we have $(AB)^T = B^T A^T$.