Lecture 6-4

Last time we reviewed the main facts about matrices; we now recall the connections between them and linear transformations. Given a linear transformation f from a finite-dimensional vector space V to itself and a basis $B = \{v_1, \dots, v_n\}$ of V we write each $f(v_i)$ as a linear combination $\sum_{j=1}^{n} a_{ji} v_j$ and then call the matrix A with jith entry a_{ji} the matrix of A relative to B. If B is replaced by a different basis $B' = \{v'_1, \dots, v'_n\}$ and if each v'_i is written as $\sum_{j=1}^{n} p_{ji} v_j$, then the matrix P with jith entry p_{ji} is invertible and the matrix of f relative to B' is $A' = P^{-1}AP$, a matrix similar to A. Given f we naturally try to find a basis B such that the matrix of f with respect to B is as simple as possible. If there is a basis B whose ith vector v_i is an eigenvector of f with eigenvalue λ_i then the matrix of f relative to B is diagonal with ith diagonal entry λ_i; we call f or A diagonalizable in this situation. More precisely, if the ith column of the matrix P is an eigenvector of A with eigenvalue λ_i then $D = P^{-1}AP$ is diagonal with ith entry λ_i along the diagonal. In general, any square matrix A has at least one (possibly complex) eigenvalue λ, since its characteristic polynomial $p(x) = \det(A - xI)$ factors into linear factors over $\mathbb{C}$. Then A is diagonalizable if and only if the geometric multiplicity of any of its eigenvalues λ (i.e. the dimension of the λ-eigenspace) matches the algebraic multiplicity (the largest power k of $x - \lambda$ dividing $p(x)$. Similar matrices have the same eigenvalues with the same algebraic and geometric multiplicities, though they need not have the same eigenvectors.

Not all matrices are diagonalizable, but we have seen a large class of diagonalizable matrices, namely the (real) symmetric $n \times n$ matrices. Any such matrix A takes the form $U^T DU = U^{-1} DU$ for some orthogonal matrix U (so that $U^T = U^{-1}$) and real diagonal matrix D; in particular, all eigenvalues of A are real. We say that A is positive definite if all eigenvalues of A are positive, or equivalently $v^T Av > 0$ for all nonzero $v \in \mathbb{R}^n$. More generally, A is said to have signature (p, q) if it has p positive eigenvalues and q negative ones (counting multiplicities); note that the multiplicity of 0 as an eigenvalue is just n minus the rank of A, so is 0 if and only if A is invertible. If A has signature (p, q) then A is congruent, but not necessarily similar, to a diagonal matrix D having p 1s and q -1s along the diagonal and all other entries 0; that is, we have $A = P^T DP$ for some invertible but not necessarily orthogonal matrix P. The quadratic form corresponding to A, that is, the homogeneous quadratic polynomial sending the row vector $x = (x_1, \ldots, x_n)$ to xAx^T, is then the sum of the squares of p linear polynomials minus the sum of the squares of q linear polynomials, where the linear polynomials involved can be taken to correspond to orthogonal but not necessarily orthonormal vectors in $\mathbb{R}^n$.

The determinantal criterion for positive definiteness asserts that a symmetric $n \times n$ matrix A is positive definite if and only if all upper left $i \times i$ minors $A^{(i)}$ of A (consisting of the entries in its first i rows and columns) have positive determinant (for $1 \leq i \leq n$). The corresponding criterion for negative definiteness is that $\det A^{(i)}$ is positive for even i but negative for odd i. Symmetric positive semidefinite matrices have unique symmetric positive semidefinite square roots; these arise in the singular value decomposition, which we will review tomorrow.

Given an inconsistent system $Ax = b$ we can replace it by its normal equations, which form the system $A^T Ax = A^T b$. This last system is always consistent and has a unique solution whenever the matrix A has full rank (i.e. rank equal to the minimum of the number of its rows and its columns). In terms of linear transformations, the transformation f with matrix A (relative to the standard bases of $\mathbb{R}^n$ and $\mathbb{R}^m$) is one-to-one when restricted to the row space of A and maps this space onto its column space, while sending the orthogonal complement of the row space of A to 0. In solving the normal equations $A^T Ax = A^T b$ we are in effect first projecting b orthogonally to the column space of A, obtaining a vector b', and then finding the unique shortest vector x with $Ax = b'$. In turn we can find x by locating any vector x' with $Ax' = b'$ and then projecting x' to the row space of A. In general, given a subspace S of $\mathbb{R}^n$ with dimension m, its orthogonal complement $S^\perp$ (consisting of all vectors orthogonal to every vector in S) has dimension $n - m$ and every $v \in \mathbb{R}^n$ is uniquely the sum of some $s \in S$ and $s' \in S^\perp$.