Lecture 5-20

We now review for the midterm on Friday. We have seen that given any linear transformation $f : V \to W$ from one finite-dimensional vector space V to another one W and a choice of bases B, B' for V, W, respectively, we get a unique matrix A of f relative to B and B', obtained as follows: if the basis B is $\{b_1, \ldots, b_n\}$ and B' is $\{b'_1, \ldots, b'_m\}$, then each $f(b_i)$ is a linear combination of b'_j, say $\sum_{i=1}^m a_{ij} b'_j$. Then a_{ij} is the ij-th entry of A, so that the ith column of A consists of the coefficients appearing when $f(b_i)$ is written as a linear combination of the b'_j. In class we have mostly concentrated on the case where $V = W$ and $b_i = b'_i$ for all i, so that the bases B and B' coincide; but you should be aware of the more general setup above and be prepared to compute with it. We have seen that two $n \times n$ matrices A, B are similar if and only if they represent the same linear transformation relative to two different bases, or if and only if $B = P^{-1}AP$ for some invertible $n \times n$ matrix P. The corresponding criterion for two $m \times n$ matrices A, B to represent the same linear transformation, each relative to its own bases of both V and W, is that $B = PAQ$, where P is an invertible $m \times m$ matrix and Q is an invertible $n \times n$ matrix.

Returning to the case of a linear transformation $f : \mathbb{R}^n \to \mathbb{R}^n$, say with matrix A relative to the standard basis of $\mathbb{R}^n$, it would be desirable to find a basis B of $\mathbb{R}^n$ that is well adapted to f (rather than just using the standard basis of $\mathbb{R}^n$). More precisely, if we can find such a basis $B = v_1, \ldots, v_n$ consisting of eigenvectors of f, say with respective eigenvalues $\lambda_1, \ldots, \lambda_n$, then the matrix of f with respect to this basis would be $D = \begin{pmatrix} \lambda_1 & & & \\ & \lambda_2 & & \\ & & \ddots & \\ & & & \lambda_n \end{pmatrix}$, a diagonal matrix; we call either f or A diagonalizable in this case. We then have $P^{-1}AP = D$, where P is the matrix whose i-th column is v_i (this matrix is guaranteed to be invertible); we also say that A is similar to D. In general, similar matrices have the same eigenvalues with the same multiplicities (both algebraic and geometric), and thus the same trace (the sum of the eigenvalues) and determinant (their product), but *not* the same eigenvectors. A matrix A is diagonalizable if and only if all of its eigenvalues a have the same geometric multiplicity (dimension of the a-eigenspace) and algebraic multiplicity (largest k for which $(a-\lambda)^k$ divides the characteristic polynomial $p(\lambda) = \det(A - \lambda I)$ of A); note that in general the roots of the characteristic polynomial are exactly the eigenvalues of A, each having geometric multiplicity at least 1.

Although many square matrices are not diagonalizable, we have learned that every real *symmetric* $n \times n$ matrix A is diagonalizable, in fact by an orthogonal matrix U, so that $U^T AU = U^{-1}AU = D$, a real diagonal matrix. More generally, the same is true of any complex *Hermitian* matrix A, that is, one for which $\bar{A}^T = A$, with U replaced by a unitary matrix (for which $U^{-1} = \bar{U}^T$). The matrix A is positive definite if and only $v^T Av > 0$ for every nonzero $v \in \mathbb{R}^n$ (or $\bar{v}^T Av > 0$ for all nonzero $v \in \mathbb{C}^n$, if A is Hermitian), where v is written as a column vector; in turn this happens if and only if all eigenvalues of A are positive. If A is real and symmetric, then it is also true that A is positive definite if and only if all of its pivots are positive, provided that every row operation on A performed in order to identify these pivots is followed immediately by the corresponding column operation, so that A remains symmetric throughout. Thanks to the preservation not only of the determinant $\det A$ of A itself, but also of the determinant $\det A^{(i)}$ of the upper left $i \times i$ submatrix of A for all $i \leq n$, whenever a multiple of a higher row of A is added to a lower one, we can furthermore say that A is positive definite if and only if $\det A^{(i)} > 0$ for all $i \leq n$. A is negative definite if and only if $\det A^{(i)}$ is negative for i odd, but positive for i even. Assuming that $\det A^{(i)} \neq 0$ for any i, any other pattern of signs among the $\det A^{(i)}$ implies that A is indefinite, having both positive and negative eigenvalues. These criteria for positive and negative definiteness and indefiniteness apply equally well to Hermitian matrices A.

Dot products and orthogonality, which play a crucial role in proving the results of the last paragraph, are also crucial to producing the best possible approximate solutions (in the sense of least squares) to inconsistent linear systems $AX = B$. Given any such system, its normal equations are those of the system $A^T AX = A^T B$, which is always consistent. If the matrix A has full rank, the normal equations will have a unique solution X, which will uniquely minimize the square length $||AX - B||^2$. In general, the normal equations will have a unique *shortest* solution X, again minimizing the square length $||AX - B||^2$. A common application in which the matrix A always has full rank is least squares polynomial fitting to data points; given any $n+1$ data points $(x_1, y_1), \ldots, (x_{n+1}, y_{n+1})$ with $x_i \neq x_j$ for $i \neq j$, and a degree $d \leq n$, there is a unique polynomial p_d of degree at most d which gives the best possible fit to the data points, so that the sum $\sum_{i=1}^{n+1} (p_d(x_i) - y_i)^2$ is minimized. The coefficient matrix is a Vandermonde matrix and always has full rank.

Whenever the square matrix A is diagonalizable, so that $A = P^{-1}DP$ with D diagonal, it is easy to compute arbitrary integral powers of A, since $A^k = P^{-1}D^kP$. In fact, in the special case where the diagonal entries of D are real and positive, we can use the equation $A^\alpha = P^{-1}D^\alpha P$ to define arbitrary real powers A^α of A, where D^α is computed by raising each of its diagonal entries to the power α. In particular, positive definite matrices always have uniquely defined square roots that are also positive definite.