Lecture 6-1

Continuing from last time, let A be an $m \times n$ complex matrix and A^* its conjugate transpose. We saw last time that the $n \times n$matrix $B = A^*A$ is Hermitian and positive semidefinite, whence there is an orthonormal basis $u_1, \ldots, u_n$ of $\mathbb{C}^n$ consisting of eigenvectors of B, say with respective eigenvalues $\lambda_1, \ldots, \lambda_n$, where the λ_i are real and nonnegative; assume that these are arranged so that $\lambda_i \neq 0$ exactly for $i \leq r$ for some index r (possibly 0). Write $Au_i = v_i \in \mathbb{C}^m$; then $A^*v_i = \lambda_i u_i$. Since A and A^* are adjoints of each other we have $(Au_i, v_i) = (v_i, v_i) = (u_i, A^*v_i) = \lambda_i$ and $(Au_i, v_j) = (u_i, A^*v_j) = 0$ for $i \neq j$, where $(\cdot, \cdot)$ is the Hermitian inner product, whence the v_i are orthogonal vectors spanning the column space (or range) C of A. We have $(v_i, v_i) = \lambda_i$, whence $v_i = 0$ if and only if $i > r$. Since the nonzero v_i form a basis of $C \subset \mathbb{C}^m$ we have $r \leq m$. Now replace the v_i equal to 0, namely those for which $i > r$, by an orthonormal basis, now denoted $v_{r+1}, \ldots, v_m$, of the orthogonal complement $C^\perp$ of C in $\mathbb{C}^m$. The upshot is that the triple matrix product $V \Sigma W$ equals A, where W is the unitary $n \times n$ matrix sending u_i to the i-th unit coordinate vector e_i, Σ is the "diagonal" $m \times n$ matrix sending e_i to $\sqrt{\lambda_i} e_i$ for $i \leq m$, and V is the unitary $m \times m$ matrix sending $\sqrt{\lambda_i} e_i$ to v_i for $i \leq r$ and e_j to v_j for $j > r$. Along the way we see that the nonzero λ_i, which are by definition the square roots of the nonzero eigenvalues of $B = A^*A$, are also the square roots of the nonzero eigenvalues of AA^*, occurring with the same multiplicity for AA^* as for A^*A, and that the eigenvectors v_i of AA^* are orthogonal. The orthonormal basis $u_1, \ldots, u_n$ of eigenvectors of B (and thus the matrix W) is not unique, but once it has been chosen it determines the first r columns of the matrix V; its remaining columns (if any) simply fill it out to a unitary $m \times m$ matrix in an arbitrary fashion. The diagonal entries of Σ are uniquely determined up to reordering. How does the polar decomposition $A = UP$ that we saw in the last lecture fit into this picture? The matrix P sends each eigenvector u_i to $\sqrt{\lambda_i} u_i$; the matrix U sends $\sqrt{\lambda_i} u_i$ for $i \leq r$ to v_i. If the set $u_1, \ldots, u_r$ is extended arbitrarily to a basis $u_1, \ldots, u_r, w_{r+1}, \ldots, w_n$ of $\mathbb{C}^n$ (e.g. by taking $w_j = u_j$ for $j > r$) then U may be extended to a linear map from $\mathbb{C}^n$ to $\mathbb{C}^m$ by defining it arbitrarily on the w_j.

The remainder of the course will be entirely devoted to review. We began the course with criteria on the second-order partials at a critical point $\vec{a}$ of a function $f : \mathbb{R}^n \to \mathbb{R}$ to have a local minimum, a local maximum, or a saddle point at $\vec{a}$. We defined the Hessian matrix H of f at $\vec{a}$ by decreeing that its ij-th entry h_{ij} is the second-order partial $f_{ij}(\vec{a})$ of f at $\vec{a}$ with respect to the i-th and j-th variables (in either order; this is a symmetric matrix). The matrix H is positive definite if and only if $v^T H v > 0$ for all nonzero vectors $v \in \mathbb{R}^n$ (by definition), or if and only if all eigenvalues of H are positive, or if and only if all square submatrices of H consisting of the entries in its first i rows and columns have positive determinant. Much more recently we learned that any Hessian matrix, or more generally any real $n \times n$ symmetric matrix A, admits an orthonormal basis of eigenvectors, each with real eigenvalue. In general such a matrix need not be either positive or negative definite, but it always has a well-defined signature (p, q), where p and q are the largest dimensions of subspaces P, Q of $\mathbb{R}^n$ such that $v^T A v > 0$ for all nonzero $v \in P, w^T A w < 0$ for all nonzero $w \in Q$. The rank of A is then $p + q$. The same results hold if A is replaced by an $n \times n$ complex Hermitian matrix (one with $\bar{A}^T = A$).

We learned the definition of differentiability for a function $\vec{f} : \mathbb{R}^n \to \mathbb{R}^m$ at a point $\vec{a} \in \mathbb{R}^n$ already last quarter, but it certainly bears repeating. Making full use of the language of linear algebra, we can say that $\vec{f}$ is differentiable at $\vec{a}$ if and only if there is an $m \times n$ matrix A such that $\lim_{\vec{h} \to \vec{0}} \frac{||\vec{f}(\vec{a}+\vec{h}) - \vec{f}(\vec{a}) - A\vec{h}||}{||\vec{h}||} = 0$; if so then the matrix A is uniquely determined and equals the Jacobian matrix J of $\vec{f} = (f_1, \ldots, f_n)$ at $\vec{a}$, whose ij-th entry a_{ij} is $\partial f_i / \partial x_j(\vec{a})$. The existence of J is *not* enough to force differentiability of $\vec{f}$ at $\vec{a}$ in general. For example, the function f defined by $f(x,y) = \frac{x^3+y^3}{x^2+y^2}$ for $(x,y) \neq (0,0)$ and $f(0,0) = 0$ fails to be differentiable at $(0,0)$; its x- and y-partials both exist and are equal to 1 at this point , but $\frac{f(h,k)-h-k}{\sqrt{h^2+k^2}}$ fails to have the limit 0 (or any limit) as $(h,k) \to (0,0)$. If the partials of the f_i exist in a neighborhood of $\vec{a}$ *and* are continuous at $\vec{a}$, then $\vec{f}$ is differentiable there. If the second-order partials $\partial^2 f_i / \partial x_j \partial x_k$ exist at $\vec{a}$ and are continuous there, then the mixed partials $\partial^2 f_i / \partial x_j \partial x_k$ and $\partial^2 f_i / \partial x_k \partial x_j$ are equal at $\vec{a}$.

If the differentiable function $f : \mathbb{R}^n \to \mathbb{R}$ is restricted to a level set of some differentiable function $g : \mathbb{R}^n \to \mathbb{R}$, then the criterion for f to have a critical point at some point $\vec{a}$ in this level set is that either $\nabla f(\vec{a}) = \lambda \nabla g(\vec{a})$ or $\nabla g(\vec{a}) = \vec{0}$. In the former case I have said that there is no second derivative test for the local nature of a critical point. Actually this was a white lie; see the Wikipedia article on Hessian matrices to learn about a determinantal criterion for testing critical points of functions on level sets individually, involving something called the "bordered Hessian matrix". You won't need to know anything about this for the final.