Lecture 3-30

Welcome back, if only virtually! Note first and foremost that the first HW assignment is *not due until Monday, April 6*, in accordance with the Provost's instructions not to assign any written work during the first week of classes.

We begin with some basic matrix definitions that come into play in the definition we gave last term of differentiability of vector-valued functions of a vector variable. A *matrix* M is a finite rectangular array of numbers; it is said to be $m \times n$ if it has m rows and n columns. The numbers (called *entries*) in an $m \times n$ matrix A are typically denoted a_{ij}, so that one uses the small letter corresponding to the capital letter denoting the matrix and denotes the entries themselves by double subscripts; here the first subscript i refers to the ith row (counting form the top) and the second one j refers to the jth column (counting from the left). Thus the upper leftmost entry of A is denoted a_{11}, the lower rightmost one a_{mn}, and so on. We can specify a matrix A by giving a formula for a_{ij} for all i, j (and making it clear what the ranges of the indices i, j are); whenever we give a formula for a_{ij}, we write A for the matrix with ijth entry a_{ij}. A column vector is an $n \times 1$ matrix, where the vector lies in $\mathbb{R}^n$. The product AV of an $m \times n$ matrix A and an $r \times 1$ column vector V is defined if and only if $r = n$; in that case it is an $m \times 1$ column vector whose ith entry from the top is $A_i \cdot V$, where A_i is the ith row of A. More generally, given two matrices A, B of possibly different sizes, the product AB is defined if and only if AB_i is defined for every column B_i of B, and in that case the ith column of AB is the column vector AB_i. Thus AB is defined if and only if A is $m \times n$ while B is $n \times r$, for some m, n, r.

We have already observed that this definition fits perfectly into the definition of differentiability for vector-valued functions. Thus given a function $\vec{f} : \mathbb{R}^n \to \mathbb{R}^m$ it is differentiable at $\vec{a} \in \mathbb{R}^n$ if and only if all coordinate functions f_i of $\vec{f}$ are differentiable at $\vec{a}$. If this holds then all gradients $\nabla f_i(\vec{a})$ must be defined. The matrix whose ith row if $\nabla f_i(\vec{a})$ (so that its ijth entry is $\partial f_i / \partial x_j(\vec{a})$, where the variables are denoted $x_1, \ldots, x_n$ as usual) is called the *Jacobian matrix of* $\vec{f}$ *at* $\vec{a}$ and is sometimes denoted $D\vec{f}(\vec{a})$, or just $D\vec{f}$, if $\vec{a}$ is understood. Now the multivariable chain rule can be stated in an especially elegant form: *if* $\vec{f} : \mathbb{R}^n \to \mathbb{R}^m$ *is differentiable at* $\vec{g}(\vec{a})$ *and* $\vec{g} : \mathbb{R}^p \to \mathbb{R}^n$ *is differentiable at* $\vec{a} \in \mathbb{R}^p$*, then the composite function* $\vec{f}(\vec{g})$ *is differentiable at* $\vec{a}$ *and* $D\vec{f}(\vec{g})(\vec{a}) = D\vec{f}(\vec{g}(\vec{a}))D\vec{g}(\vec{a})$. This holds because if we write $f_1, \ldots, f_m$ for the coordinates of f, regard each f_i as a function of the coordinates $g_1, \ldots, g_n$ of $\vec{g}$, and finally denote the variables on which $\vec{g}$ depends as $x_1, \ldots, x_p$, then we have $\partial f_i / \partial x_k = \sum_{j=1}^{n} (\partial f_i / \partial g_j)(\partial g_j / \partial x_k)$. Here of course it must be understood that the partials of the f_i are to be evaluated at $\vec{g}(\vec{a})$, while the partials of the g_j are to be evaluated at $\vec{a}$.

With this piece of linear algebra under our belt, we now return to Chapter 16 of Salas-Hille. We left off with real-valued functions f defined on an open subset U of $\mathbb{R}^n$; following our earlier work with real-valued functions of a real variable, we want to give conditions for f to have a local maximum or minimum at $\vec{a} \in U$. We have already observed (and it is easy to see directly) that in order for $\vec{a}$ to have a chance of being a local maximum or minimum of f, we must have $(\partial f / \partial x_i)(\vec{a})$ either undefined or equal to 0 for all indices i; accordingly, we say that $\vec{a} \in U$ is *critical* (or *stationary*) for f if this condition holds. (In practice, we will ignore the possibility that some partial of f is undefined at $\vec{a}$, concentrating on the case where $\nabla f(\vec{a}) = \vec{0}$.) Now a natural question is whether a critical point $\vec{a}$ for f is a local maximum, a local minimum, or neither (if it is neither, it is called a *saddle point*, as in the one-variable case). To answer this question we have to look at the quadratic rather than just the linear approximation of f at $\vec{a} = (a, b)$. If $n = 2$ and A, B, C respectively denote $f_{xx}(\vec{a}), f_{xy}(\vec{a}), f_{yy}(\vec{a})$ then, assuming f is twice differentiable at $\vec{a}$, we get that $\frac{f(a+h,b+k)-f(a,b)-(A/2)h^2-Bhk-(C/2)k^2}{h^2+k^2} \to 0$ as $(h, k) \to (0, 0)$ from the definition of twice differentiability. This reduces matters to studying the homogeneous quadratic polynomial $q(h, k) = (A/2)h^2 + Bhk + (C/2)k^2$; it is not difficult to show that if $q(h, k) > 0$ for all $(h, k) \neq (0, 0)$, then $q(h, k)/(h^2 + k^2)$ is bounded between two positive constants for all $(h, k) \neq (0, 0)$, whence $f(a + h, b + k) > f(a, b)$ for all such (h, k) and f has a local minimum at (a, b). Conversely, if instead $q(h, k) < 0$ for all $(h, k) \neq (0, 0)$, then $f(a + h, b + k) < f(a, b)$ for all such (h, k) and f has a local maximum at (a, b). Now we can also have $q(h, k) > p(h^2 + k^2)$ whenever (h, k) is a nonzero multiple of one vector $\vec{v}$, for some positive constant p, while $q(h, k), q(h^2 + k^2)$ whenever (h, k) is a nonzero multiple of another vector $\vec{w}$, for some negative constant q. In this case we can say for sure that (a, b) is a saddle point for f. Finally, it is possible that $q(h, k) = 0$ for some $(h, k) \neq (0, 0)$; in this case, $q(h, k)$ does not give us enough information to determine whether f has a local maximum, a local minimum, or a saddle point at (a, b).

Thus we must decide for a given homogeneous quadratic function $q(x, y) = ax^2 + bxy + cy^2$ when it takes only positive values, when only negative values, and when both positive and negative values, for x, y not both 0. Dividing by x^2 and regarding the resulting function as a quadratic polynomial in y/x it is easy to decide that $q(x, y)$ takes only positive values (we say it is *positive definite*) if and only if $a > 0, b^2 - 4ac < 0$; it takes only negative values (and is called *negative definite*) if and only if $a < 0, b^2 - 4ac < 0$; and finally it takes both positive and negative values (and is called *indefinite*) if $b^2 - 4ac > 0$. If $b^2 - 4ac = 0$ then $q(x, y)$ takes the value 0 infinitely often and all bets are off. Translating back to our function f and recalling our earlier notation, we find that *if $A > 0, AC - B^2 > 0$ then f has a local minimum at $\vec{a}$; if $A < 0, AC - B^2 > 0$, then f has a local maximum at $\vec{a}$; if $AC - B^2 < 0$, then f has a saddle point at $\vec{a}$.* The case $AC - B^2 = 0$ is indeterminate; any behavior of $\vec{a}$ as a critical point is possible in this case.