Lecture 3-3

Continuing from last time, let $f : \mathbb{R}^n \to \mathbb{R}$ be differentiable at a point $\vec{a} \in \mathbb{R}^n$. One way of measuring the rate of change of f at $\vec{a}$ while avoiding the pitfall of dividing by a vector is to focus on a particular direction, given by a unit vector $\vec{u} \in \mathbb{R}^n$, and compute $\lim_{t\to 0} \frac{f(\vec{a}+t\vec{u})-f(\vec{a})}{t}$; we call this quantity the *directional derivative of* f *in the* $\vec{u}$ *direction at* $\vec{a}$ (but there is no standard notation for it). This is easily computed in terms of $\nabla f(\vec{a})$. Indeed, we have $\lim_{t\to 0} \frac{f(\vec{a}+t\vec{u})-f(\vec{a})-t\vec{u}\cdot\nabla f(\vec{a})}{|t|} = 0$, since $|t| = ||t\vec{u}||$, from which it easily follows that the directional derivative of f in the $\vec{u}$ direction at $\vec{a}$ is given by the dot product $\nabla f(\vec{a}) \cdot \vec{u}$. In particular, invoking an earlier calculation, we find that *if* $\nabla f(\vec{a}) \neq \vec{0}$, *then the direction of* $\nabla f(\vec{a})$ *is the direction of maximal rate of increase of* f *at* $\vec{a}$ *and the length of* $\nabla f(\vec{a})$ *is the rate of increase of* f *in this direction.* This gives a direct description of $\nabla f(\vec{a})$ without reference to formulas or coordinates.

Much more generally, suppose again that f is differentiable by $\vec{a}$ and now that $\vec{r}(t)$ is a parametrized curve passing through $\vec{a}$ at time $t = t_0$. Then we have the composite function sending t to $f(\vec{r}(t))$ and it is natural to ask what its derivative is at time t_0. The answer is given by the multivariable analogue of the chain rule (unsurprisingly given the same name): *the derivative* $f(\vec{r}(t_0))$ *of* $f(\vec{r}(t))$ *at* t_0 *is given by* $\nabla f(\vec{a}) \cdot \vec{r}'(t_0)$ *(or, in words, by the dot product of the gradient and tangent vectors of* f *and* $\vec{r}$ *at* $\vec{a}$*, respectively).* To see this we again just evaluate the limit giving this derivative directly: we have that $\frac{f(\vec{r}(t_0+h))-f(\vec{r}(t_0))-\nabla f(\vec{a})\cdot(\vec{r}(t_0+h)-\vec{r}(t_0))}{||\vec{r}(t_0+h)-\vec{r}(t_0)}\to 0$ as h goes to 0, avoiding nonzero values for h such that $\vec{r}(t_0 + h) = \vec{r}(t_0)$, for which the fraction is undefined. Multiplying by $||\frac{\vec{r}(t_0+h)-\vec{r}(t_0)}{h}||$, which approaches $||\vec{r}'(t_0)||$ as $h \to 0$, we see that $\lim_{h\to 0} \frac{f(\vec{r}(t_0+h)-\vec{r}(t_0)-\nabla f(\vec{a})\cdot(\vec{r}(t_0+h)-\vec{r}(t_0)}{h} = 0$; clearly this last limit still holds even if nonzero h's with $\vec{r}(t_0 + h) = \vec{r}(t_0)$ are allowed. Since $\frac{\vec{r}(t_0+h)-\vec{r}(t_0)}{h} \to \vec{r}'(t_0)$ as $h \to 0$, the desired result follows.

The rest of this week will be spent in working out consequences of the chain rule; it is the single most useful result in multivariable differential calculus. In terms of partial derivatives, and changing notation a bit, the chain rule says that *if a real-valued function* f *depends differentiably on variables* $x_1, \ldots, x_n$*, which in turn depend differentiably on* t*, then* $\frac{df}{dt} = \sum_{i=1}^{n} \frac{\partial f}{\partial x_i}\frac{dx_i}{dt}$; notice that we use the notation d rather than ∂ for the derivatives of f and the x_i with respect to t since these functions depend on only one variable. Notice also that we (shamefully but, alas, standardly) use the same notation f for the composite function of t and the multivariable function of the x_i. Notice further that the right-hand side is given to us as a sum of products, but can also be viewed as a single (dot) product. Notice finally, that the same result holds, replacing all d's by ∂'s, if in fact the x_i depend on other variables besides t, since such variables are held constant in computing $\partial f/\partial t$. One must be especially careful about notation in this last case, however, since if only ∂'s and no d's occur in a derivative formula then there will be no typographical clue about whether one is referring to a composite or a noncomposite function.

We can now return to and justify an earlier computation in differential equations. Suppose that the implicit equation $f(x_1, \ldots, x_n) = c$ for constant c is satisfied whenever x_n equals a given function $g(x_1, \ldots, x_{n-1})$ of the other variables $x_1, \ldots, x_{n-1}$, where f and g are both differentiable. By the chain rule, on differentiating the formula $f(x_1, \ldots, x_n) =$

c with respect to x_n we get $(\partial f/\partial x_i) + (\partial f/\partial x_n)(\partial x_n/\partial x_i) = 0$, whence $\partial x_n/\partial x_i = \partial g/\partial x_i = \frac{-\partial f/\partial x_i}{\partial f/\partial x_n}$ at any point where $\partial f/\partial x_n \neq 0$. (Note that if we just mindlessly cancelled the ∂f here, we would have been led to the *false* formula $\frac{\partial f/\partial x_i}{\partial f/\partial x_n}$; the correct formula definitely involves the minus sign.) In fact, we do not need to assume that x_n is a differentiable function of the other variables in this setting; a deep result called the Implicit Function Theorem states that the equation $f(x_1, \dots, x_n) = c$ can *always* be solved uniquely and differentiably for x_n in terms of the other variables near a point $\vec{a}$ with $f(\vec{a}) = c$, provided that $\partial f/\partial x_n(\vec{a}) \neq 0$, and in this case we have $\partial x_n/\partial x_i(b_1, \dots, b_{n-1}) = -\partial f/\partial x_i(b_1, \dots, b_n)/\partial f/\partial x_n(b_1, \dots, b_n)$ for all $\vec{b} = (b_1, \dots, b_n)$ sufficiently close to $\vec{a}$ with $f(\vec{b}) = c$.

In particular, the general solution to an exact differential equation $M\,dx + N\,dy = 0$ is indeed $f(x, y) = c$ with a c a constant, if f is chosen so that $\partial f/\partial x = M, \partial f/\partial y = N$.

Combining the chain rule with the mean value theorem, we get a version of the mean value theorem that holds for functions of several variables: if a function $f : \mathbb{R}^n \to \mathbb{R}$ is defined and differentiable on the line segment from $\vec{a}$ to $\vec{b}$ with $\vec{a}, \vec{b} \in \mathbb{R}^n$, then we have $f(\vec{b}) - f(\vec{a}) = (\vec{b} - \vec{a}) \cdot \nabla f(\vec{c})$ for some $\vec{v}$ strictly between $\vec{a}$ and $\vec{b}$ and on the line segment joining them. To see this, applying the mean value theorem to the function $g : [0, 1] \to \mathbb{R}$ defined by $g(t) = f((1 - t)\vec{a} + t\vec{b})$. In particular, if f is differentiable on a convex subset C of $\mathbb{R}^n$ (one containing the line segment joining any two of its points) and if $\nabla f = \vec{0}$ on C, then f is constant on C. In fact, this holds more generally if any two points of C can be joined by a parametrized curve lying entirely in C.