Lecture 6: 2.5 The Chain Rule.

The **Chain rule in one variable**: suppose that $y = g(x)$, and $z = f(y)$, i.e. $z = h(x)$, where $h(x) = f(g(x)) = f \circ g(x)$ then

$$\frac{dz}{dx} = \frac{dz}{dy}\frac{dy}{dx} \qquad \Leftrightarrow \qquad h'(x) = f'(y)\,g'(x)$$

The intuitive way to understand this is through the linear approximation:

$$\triangle z = f(y + \triangle y) - f(y) \sim f'(y)\triangle y$$

which is the same as saying that f is differentiable. Similarly

$$\triangle y = g(x + \triangle x) - g(x) \sim g'(x)\triangle x$$

and if we combine the two we get

$$\triangle z \sim f'(y)\,g'(x)\triangle x$$

Since this is also equal to

$$\triangle z = h(x + \triangle x) - h(x) \sim h'(x)\triangle x$$

the chain rule in one variable follows.

The **Chain rule in several variables**: Suppose that $g : \mathbf{R}^n \to \mathbf{R}^m$, $f : \mathbf{R}^m \to \mathbf{R}^p$ and let $h = f \circ g : \mathbf{R}^n \to \mathbf{R}^p$ (i.e. $h(x) = f(g(x))$). Then

$$Dh(x_0) = Df(y_0)\,Dg(x_0), \qquad \text{where} \qquad y_0 = g(x_0)$$

and the right hand side is the $p \times n$ matrix formed by the matrix product of the $p \times m$ matrix $Df(y_0)$ by the $m \times n$ matrix $Dg(x_0)$.

The intuitive argument above actually generalizes to several variables just by replacing f' by Df etc. since differentiability of functions in several variables says

$$g(x + \triangle x) - g(x) \sim Dg(x)\triangle x.$$

If $h(t) = f(\mathbf{c}(t))$ where $f : \mathbf{R}^3 \to \mathbf{R}$, $\mathbf{c}(t) = (x(t), y(t), z(t))$ is a path or curve, then by the chain rule

$$\frac{\partial h}{\partial t} = Df\,D\mathbf{c} = \begin{bmatrix} \dfrac{\partial f}{\partial x} & \dfrac{\partial f}{\partial y} & \dfrac{\partial f}{\partial z} \end{bmatrix} \begin{bmatrix} \dfrac{dx}{dt} \\[2mm] \dfrac{dy}{dt} \\[2mm] \dfrac{dz}{dt} \end{bmatrix} = \frac{\partial f}{\partial x}\frac{dx}{dt} + \frac{\partial f}{\partial y}\frac{dy}{dt} + \frac{\partial f}{\partial z}\frac{dz}{dt}$$

The **gradient** of a function $f : \mathbf{R}^n \to \mathbf{R}$ given by

$$\mathbf{grad}\,f = \nabla f = \begin{bmatrix} \dfrac{\partial f}{\partial x_1} & \cdots & \dfrac{\partial f}{\partial x_n} \end{bmatrix}$$

1

This can also be expressed with the gradient notation and dot product

$$\frac{dh}{dt}(t) = \nabla f(\mathbf{c}(t)) \cdot \mathbf{c}'(t)$$

where

$$\mathbf{c}'(t) = (x'(t), y'(t), z'(t)),$$

is the velocity vector of the path.

Note that $\mathbf{c}'(t)$ is tangent to the path $\mathbf{c}(t)$, which follows since it can be obtained as the limit as $h \to 0$ of the vector between two close points on the curve

$$\frac{\mathbf{c}(t+h) - \mathbf{c}(t)}{h} = \left(\frac{x(t+h) - x(t)}{h}, \frac{y(t+h) - y(t)}{h}, \frac{z(t+h) - z(t)}{h}\right) \to (x'(t), y'(t), z'(t))$$

Ex. If $z = x^2 + y^2$, $x = \cos t$ and $y = \sin t$ find dz/dt.

Sol. 1 $\quad \dfrac{dz}{dt} = \dfrac{\partial z}{\partial x}\dfrac{dx}{dt} + \dfrac{\partial z}{\partial y}\dfrac{dy}{dt} = 2x\,(-\sin t) + 2y\,\cos t = 2\cos t\,(-\sin t) + 2\sin t\cos t = 0$

Sol. 2 $\qquad z = x^2 + y^2 = \cos^2 t + \sin^2 t = 1, \qquad \dfrac{dz}{dt} = 0$

Ex. If $z = h(r, \theta) = f(x, y)$ where $x = g_1(r, \theta) = r\cos\theta$ and $y = g_2(r, \theta) = r\sin\theta$ then by the chain rule

$$\left(\frac{\partial h}{\partial r}, \frac{\partial h}{\partial \theta}\right) = \left(\frac{\partial f}{\partial x}, \frac{\partial f}{\partial y}\right) \begin{pmatrix} \dfrac{\partial g_1}{\partial r} & \dfrac{\partial g_1}{\partial \theta} \\ \dfrac{\partial g_2}{\partial r} & \dfrac{\partial g_2}{\partial \theta} \end{pmatrix}$$

i.e.

$$\frac{\partial h}{\partial r} = \frac{\partial f}{\partial x}\frac{\partial x}{\partial r} + \frac{\partial f}{\partial y}\frac{\partial y}{\partial r}$$

$$\frac{\partial h}{\partial \theta} = \frac{\partial f}{\partial x}\frac{\partial x}{\partial \theta} + \frac{\partial f}{\partial y}\frac{\partial y}{\partial \theta}$$

or written shorter since h and f is the same function expressed in different coordinates

$$\frac{\partial}{\partial r} = \frac{\partial x}{\partial r}\frac{\partial}{\partial x} + \frac{\partial y}{\partial r}\frac{\partial}{\partial y}$$

$$\frac{\partial}{\partial \theta} = \frac{\partial x}{\partial \theta}\frac{\partial}{\partial x} + \frac{\partial y}{\partial \theta}\frac{\partial}{\partial y}$$

In this case the matrix

$$\begin{pmatrix} \dfrac{\partial x}{\partial r} & \dfrac{\partial x}{\partial \theta} \\ \dfrac{\partial y}{\partial r} & \dfrac{\partial y}{\partial \theta} \end{pmatrix} = \begin{pmatrix} \cos\theta & -r\sin\theta \\ \sin\theta & r\cos\theta \end{pmatrix}$$

Note that in general $\dfrac{\partial r}{\partial x} \neq \left(\dfrac{\partial x}{\partial r}\right)^{-1}$. whereas for a function of one variable its true that $dx/dy = (dy/x)^{-1}$. That this is true in one dimension follows from differentiating the identity $f(f^{-1}(x)) = x$ which gives $f'(f^{-1}(x))f^{-1}(x)' = 1$. The higher dimension analogue of this would be with the derivative matrices so

$$\begin{pmatrix} \dfrac{\partial r}{\partial x} & \dfrac{\partial r}{\partial y} \\ \dfrac{\partial \theta}{\partial x} & \dfrac{\partial \theta}{\partial y} \end{pmatrix} = \begin{pmatrix} \dfrac{\partial x}{\partial r} & \dfrac{\partial x}{\partial \theta} \\ \dfrac{\partial y}{\partial r} & \dfrac{\partial y}{\partial \theta} \end{pmatrix}^{-1} = \frac{1}{r}\begin{pmatrix} r\cos\theta & r\sin\theta \\ -\sin\theta & -\cos\theta \end{pmatrix}$$

2.6 The gradient and directional derivative.

The **gradient** of a function $f : \mathbf{R}^3 \to \mathbf{R}$ is the vector

$$\nabla f = \left(\frac{\partial f}{\partial x}, \frac{\partial f}{\partial y}, \frac{\partial f}{\partial z} \right)$$

i.e. it is the matrix of derivatives written as a vector.

Consider the equation of a line in space $\ell(t) = \mathbf{x} + t\mathbf{v}$, $-\infty < t < \infty$.
The function $h(t) = f \circ \ell(t) = f(\mathbf{x} + t\mathbf{v})$ represents the function f restricted to the line.
The **directional derivative** of f at $\mathbf{x}$ in the direction of unit vector $\mathbf{v}$ is given by

$$\frac{d}{dt} f(\mathbf{x} + t\mathbf{v}) \Big|_{t=0} = \nabla f(\mathbf{x}) \cdot \mathbf{v}$$

Here the equality follows from the chain rule: $h' = Dh = Df \, D\ell = \nabla f \cdot \ell' = \nabla f \cdot \mathbf{v}$.

The reason we choose $\mathbf{v}$ to be a unit vector is that we want the directional derivative to represent the rate of change in different directions.
Suppose that f represents the temperature at different points in space. Suppose that a fly flies along the line above at unit speed then the change of temperature per unit time or distance is the directional derivative.

The gradient points in the direction along which f increases the fastest.
In fact $\nabla f \cdot \mathbf{v} = |\nabla f| \, |\mathbf{v}| \cos \theta$, where θ is the angle between ∇f and $\mathbf{v}$, and the max is when $\cos \theta = 1$.

Suppose we are lost in wood and we want to reach a high hill top to see where we are. However, we can only see a few feet in front of us because of the high trees. In which direction shall we go in order to reach a hill-top fast. The answer is that if we go in the direction of the grade likely to reach a hill-top fast.

The gradient is normal to the tangent plane of the level surface:

Let $f : \mathbf{R}^3 \to \mathbf{R}$ and let (x_0, y_0, z_0) be a point on the level surface S defined by $f(x, y, z) = k$, for some constant k. Then $\nabla f(x_0, y_0, z_0)$ is normal to the level surface in the following sense: If $\mathbf{v} = \mathbf{c}'(0)$ is a tangent vector to a path $\mathbf{c}(t)$ in S with $\mathbf{c}(0) = (x_0, y_0, z_0)$, then $\nabla f(x_0, y_0, z_0) \cdot \mathbf{v} = 0$.
In fact, since $f(x(t), y(t), z(t)) = k$ it follows that

$$0 = \frac{d}{dt} f(x(t), y(t), z(t)) = \nabla f(x(t), y(t), z(t)) \cdot \mathbf{c}'(t)$$

Let S be a level surface $f(x, y, z) = k$. The **tangent plane** of S at a point (x_0, y_0, z_0) of S is defined by the equation

$$\nabla f(x_0, y_0, z_0) \cdot (x - x_0, y - y_0, z - z_0) = 0$$

In fact (x, y, z) is in the tangent plane if $(x, y, z) - (x_0, y_0, z_0)$ is parallel to the plane and hence perpendicular to the normal.