Lecture 4: 2.3 Differentiation.

Given $f : \mathbf{R}^3 \to \mathbf{R}$ The **partial derivative** of f with respect x is defined by

$$f_x(x, y, z) = \frac{\partial f}{\partial x}(x, y, z) = \lim_{h \to 0} \frac{f(x+h, y, z) - f(x, y, z)}{h}$$

if it exist. The partial derivatives $\partial f / \partial y$ and $\partial f / \partial z$ are defined similarly and the extension to functions of n variables is analogous.

What is the meaning of the derivative of a function $y = f(x)$ of one variable?
It is the slope of the tangent line $y = f(x_0) + f'(x_0)(x - x_0)$ and the tangent line is that it is the linear function that best approximates $f(x)$ when x is close to x_0.

The **linear approximation** of $f(x, y)$ at (x_0, y_0) is

$$\ell_{x_0, y_0}(x, y) = f(x_0, y_0) + f_x(x_0, y_0)(x - x_0) + f_y(x_0, y_0)(y - y_0)$$

A function $f(x, y)$ is called **differentiable** at (x_0, y_0) if the partial derivatives exist and if

$$\lim_{(x,y) \to (x_0, y_0)} \frac{f(x, y) - \ell_{(x_0, y_0)}(x, y)}{\|(x, y) - (x_0, y_0)\|} = 0$$

The definition of differentiability is motivated by the idea that the **tangent plane** to the surface $z = f(x, y)$ at (x_0, y_0), i.e. the plane

$$z = \ell_{x_0, y_0}(x, y)$$

should give a good approximation to the function f close to (x_0, y_0).
If $\mathbf{F} : \mathbf{R}^3 \to \mathbf{R}^3$ is given by $\mathbf{F}(\mathbf{x}) = \big(f_1(x_1, x_2, x_3), f_2(x_1, x_2, x_3), f_3(x_1, x_2, x_3)\big)$ then the **derivative matrix** of partial derivatives is defined by

$$\mathbf{DF} = \begin{bmatrix} \dfrac{\partial f_1}{\partial x_1} & \dfrac{\partial f_1}{\partial x_2} & \dfrac{\partial f_1}{\partial x_3} \\ \dfrac{\partial f_2}{\partial x_1} & \dfrac{\partial f_2}{\partial x_2} & \dfrac{\partial f_2}{\partial x_3} \\ \dfrac{\partial f_3}{\partial x_1} & \dfrac{\partial f_3}{\partial x_2} & \dfrac{\partial f_3}{\partial x_3} \end{bmatrix} = \begin{bmatrix} \dfrac{\partial \mathbf{F}}{\partial x_1} & \dfrac{\partial \mathbf{F}}{\partial x_2} & \dfrac{\partial \mathbf{F}}{\partial x_3} \end{bmatrix}$$

Here the second expression mean that the we think of $\mathbf{F} = \begin{bmatrix} f_1 \\ f_2 \\ f_3 \end{bmatrix}$ as a column vector and $\dfrac{\partial \mathbf{F}}{\partial x_1}$ as the derivative of the column vector. Similarly we define the $m \times n$ matrix $\mathbf{DF}$ for a function $\mathbf{F} : \mathbf{R}^n \to \mathbf{R}^n$.

A special case is the **gradient** of a function $f : \mathbf{R}^n \to \mathbf{R}$ given by

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

A function $\mathbf{F}:\mathbf{R}^n \to \mathbf{R}^m$ is called differentiable at $\mathbf{a}$ if

$$\lim_{\mathbf{x}\to\mathbf{a}} \frac{\|\mathbf{F}(\mathbf{x})-\mathbf{F}(\mathbf{a})-\mathbf{DF}(\mathbf{a})(\mathbf{x}-\mathbf{a})\|}{\|\mathbf{x}-\mathbf{a}\|}=0$$

Here $\mathbf{DF}(\mathbf{a})(\mathbf{x}-\mathbf{a})$ is matrix multiplication of the $m\times n$ matrix or column vector $\mathbf{DF}(\mathbf{a})$ by the $n\times 1$ matrix $\mathbf{x}-\mathbf{a}$. If $f:\mathbf{R}^2\to\mathbf{R}$ then

$$\mathbf{D}f(\mathbf{a})(\mathbf{x}-\mathbf{a}) = \begin{bmatrix} \frac{\partial f}{\partial x_1}(\mathbf{a}) & \frac{\partial f}{\partial x_2}(\mathbf{a}) \end{bmatrix} \begin{bmatrix} x_1-a_1 \\ x_2-a_2 \end{bmatrix} = \frac{\partial f}{\partial x_1}(\mathbf{a})(x_1-a_1)+\frac{\partial f}{\partial x_2}(\mathbf{a})(x_2-a_2)$$

If $\mathbf{F}:\mathbf{R}^n\to\mathbf{R}^m$ then

$$\mathbf{DF}(\mathbf{a})(\mathbf{x}-\mathbf{a}) = \begin{bmatrix} \frac{\partial \mathbf{F}}{\partial x_1}(\mathbf{a}) & \dots & \frac{\partial \mathbf{F}}{\partial x_n}(\mathbf{a}) \end{bmatrix} \begin{bmatrix} x_1-a_1 \\ \dots \\ x_n-a_n \end{bmatrix}$$

$$= \frac{\partial \mathbf{F}}{\partial x_1}(\mathbf{a})(x_1-a_1)+\cdots+\frac{\partial \mathbf{F}}{\partial x_n}(\mathbf{a})(x_n-a_n) = \begin{bmatrix} \frac{\partial f_1}{\partial x_1} \\ \vdots \\ \frac{\partial f_m}{\partial x_1} \end{bmatrix}(x_1-a_1)+\cdots+\begin{bmatrix} \frac{\partial f_1}{\partial x_n} \\ \vdots \\ \frac{\partial f_m}{\partial x_n} \end{bmatrix}(x_n-a_n)$$

We can think of the derivative of $\mathbf{F}$ at the point $\mathbf{a}\in\mathbf{R}^n$ as the linear map $\mathbf{DF}:\mathbf{R}^n\to\mathbf{R}^m$, mapping the vector $\mathbf{h}=(h_1,...,h_n)$ to the vector

$$\mathbf{DF}(\mathbf{a})\mathbf{h} = \lim_{t\to 0}\frac{\mathbf{F}(\mathbf{a}+t\mathbf{h})-\mathbf{F}(\mathbf{a})}{t} = \frac{\partial \mathbf{F}}{\partial x_1}(\mathbf{a})h_1+...+\frac{\partial \mathbf{F}}{\partial x_n}(\mathbf{a})h_n,$$

2.4 Paths and curves.

A **path** or a curve in $\mathbf{R}^3$ is a map $\mathbf{c}:I\to\mathbf{R}^3$ of an interval $I=[a,b]$ to $\mathbf{R}^3$, i.e. for each $t\in I$ $\mathbf{c}(t)$ is a vector $\mathbf{c}=x(t)\mathbf{i}+y(t)\mathbf{j}+z(t)\mathbf{k}$.

The **velocity** of a path is defined to be

$$\mathbf{c}'(t)=\lim_{h\to 0}\frac{\mathbf{c}(t+h)-\mathbf{c}(t)}{h} = \mathbf{x}'(t)\mathbf{i}+\mathbf{y}'(t)\mathbf{j}+\mathbf{z}'(t)\mathbf{k}$$

The **tangent line** to the curve at t_0 is given by

$$\ell(t)=\mathbf{c}(t_0)+\mathbf{c}'(t_0)(t-t_0)$$

Ex Let $\mathbf{c}(t)=(\cos t,\sin t,t)$. Find the equation of the tangent line at $t=\pi/2$.
Sol We have $\mathbf{c}'(t)=(-\sin t,\cos t,1)$. The equation of the tangent line is

$$\ell(t)=\mathbf{c}(\pi/2)+\mathbf{c}'(\pi/2)(t-\pi/2)=(0,1,\pi/2)+(1,0,1)(t-\pi/2)$$

or $x=-(t-\pi/2)$, $y=1$ and $z=\pi/2+(t-\pi/2)$.

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

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