Lecture 9: 4.1 Taylor's formula in several variables.

Recall Taylors formula for $f\colon \mathbf{R} \to \mathbf{R}$:

$$(1)\quad f(x) = f(a) + f'(a)(x-a) + \frac{f''}{2}(a)(x-a)^2 + ... + \frac{f^{(k)}(a)}{k!}(x-a)^k + R_k(x-a,a)$$

where the remainder or error tends to 0 faster than the previous terms when $x \to a$:

$$(2)\qquad |R_k(x-a,a)| \le \frac{M}{(k+1)!}|x-a|^{k+1}, \qquad \text{if} \quad |f^{(k+1)}(z)| \le M,$$

for $|z-a| \le |x-a|$. The Taylor polynomial $P_k = f_k - R_k$ is the polynomial of degree k that best approximate $f(x)$ for x close to a. It is chosen so its derivatives of order $\le k$ are equal to the derivatives of f at a. (2) follows from repeated integration of

$$(2b)\qquad \frac{d^{k+1}}{dx^{k+1}} R_k(x-a,a) = f^{k+1}(x), \qquad \frac{d^j}{dx^j} R_k(x-a,a)\Big|_{x=a} = 0, \quad j \le k.$$

A similar formula hold for functions of several variables $\mathbf{F}\colon \mathbf{R}^n \to \mathbf{R}^m$. In order to state it we first write $\mathbf{x} = (x_1, x_2, ..., x_n) \in \mathbf{R}^n$ and $\mathbf{a} = (a_1, ..., a_n) \in \mathbf{R}^n$.

$$(3)\quad \mathbf{F}(\mathbf{x}) = \mathbf{F}(\mathbf{a}) + \sum_{i=1}^n \mathbf{F}_{x_i}(\mathbf{a})(x_i - a_i) + \frac{1}{2}\sum_{i,j=1}^n \mathbf{F}_{x_i\, x_j}(\mathbf{a})(x_i - a_i)(x_j - a_j)$$

$$+ ... + \frac{1}{k!}\sum_{i_1,...,i_k=1}^n \mathbf{F}_{x_{i_1}\cdots x_{i_k}}(\mathbf{a})(x_{i_1} - a_{i_1})\cdots(x_{i_k} - a_{i_k}) + \mathbf{R}_k(\mathbf{x}-\mathbf{a},\mathbf{a}),$$

where the remainder or error tends to 0 faster than the previous terms when $x \to a$:

$$(4)\quad \|\mathbf{R}_k(x,a)\| \le \frac{M}{(k+1)!}\|\mathbf{x}-\mathbf{a}\|^{k+1}, \qquad \text{if} \quad \sum_{i_1,...,i_{k+1}=1}^n \|\mathbf{F}_{x_{i_1}\cdots x_{i_{k+1}}}(\mathbf{z})\| \le M,$$

for $\|\mathbf{z}-\mathbf{a}\| \le \|\mathbf{x}-\mathbf{a}\|$. Here

$$(5)\qquad \mathbf{F}_{x_{i_1}\cdots x_{i_k}} = \frac{\partial^k \mathbf{F}}{\partial x_{i_1}\cdots \partial x_{i_k}}$$

First, the general case reduces to the case $m = 1$ by considering each component of $\mathbf{F} = (F_1, ..., F_m)$ and we may hence assume that $\mathbf{F}\colon \mathbf{R}^n \to \mathbf{R}$. In order to prove (3) we introduce $\mathbf{x} - \mathbf{a} = \mathbf{h}$ and apply the one dimensional Taylor's formula (1) to the function $f(t) = \mathbf{F}(\mathbf{x}(t))$ along the line segment $\mathbf{x}(t) = \mathbf{a} + t\mathbf{h}$, $0 \le t \le 1$:

$$(6)\qquad f(1) = f(0) + f'(0) + f''(0)/2 + ... + f^{(k)}(0)/k! + R_k$$

Here $f(1) = \mathbf{F}(\mathbf{a}+\mathbf{h})$, i.e. the left hand side of (3), $f(0) = \mathbf{F}(\mathbf{a})$, i.e. the first term in the right hand side of (3), and by the chain rule

$$(7)\qquad f'(t) = \frac{d}{dt}\mathbf{F}(\mathbf{x}(t)) = \sum_{i=1}^n \mathbf{F}_{x_i}(\mathbf{x}(t))\frac{dx_i}{dt} = \sum_{i=1}^n \mathbf{F}_{x_i}(\mathbf{x}(t))h_i$$

and hence $f'(0)$ is the second term in the right of (3). Repeating this gives

$$(8)\qquad f''(t) = \frac{d}{dt}\sum_{i=1}^n \mathbf{F}_{x_i}(\mathbf{x}(t))h_i = \sum_{i=1,j=1}^n \mathbf{F}_{x_i x_j}(\mathbf{x}(t))h_i h_j$$

and this gives the third term and so on.

If $f:\mathbf{R}^2\to\mathbf{R}$, $\mathbf{a}=(0,0)$ and $\mathbf{x}=(x,y)$ then the second degree Taylor polynomial is

$$f(x,y)\sim f(0,0)+f_x(0,0)x+f_y(0,0)y+\frac{1}{2}\big(f_{xx}(0,0)x^2+2f_{xy}(0,0)xy+f_{yy}(0,0)y^2\big)$$

Here we used the equality of mixed partial derivatives $f_{xy}=f_{yx}$.

Ex. 1 Let $f(x,y)=3+2x+x^2+2xy+3y^2+x^3-y^4$. Find the second degree Taylor polynomial around $\mathbf{a}=(0,0)$.

Sol. The second degree Taylor polynomial is

$$f(0,0)+f_x(0,0)x+f_y(0,0)y+\frac{1}{2}\big(f_{xx}(0,0)x^2+2f_{xy}(0,0)xy+f_{yy}(0,0)y^2\big)$$
$$=3+2x+\frac{1}{2}\big(2x^2+2\cdot 2xy+6y^2\big)=3+2x+x^2+2xy+3y^2$$

The derivative of a vector field as a linear map. Let $\mathbf{F}:\mathbf{R}^n\to\mathbf{R}^m$ be a vector field. Then 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 directional derivative in the direction of $\mathbf{h}$

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

This is the linear map that best approximates the function close to $\mathbf{a}$:

$$\mathbf{F}(\mathbf{a}+\mathbf{h})=\mathbf{F}(\mathbf{a})+\mathbf{DF}(\mathbf{a})\mathbf{h}+\mathbf{R}_2(\mathbf{a},\mathbf{h}),\qquad\text{where}\quad|\mathbf{R}_2(\mathbf{a},\mathbf{h})|\le M|\mathbf{h}|^2,$$

tends to 0 faster than the other terms as $|\mathbf{h}|\to 0$.

Ex.(if time) Let $\mathbf{F}(x,y)=(1+x-y+x^2)\mathbf{i}+(x^2-y^2+y^4)\mathbf{k}$. Find the Taylor polynomial of degree one for $\mathbf{F}(x,y)$ around $(x,y)=(1,0)$. **Sol.**

$$\frac{\partial\mathbf{F}}{\partial x}(x,y)=\big((1+2x)\mathbf{i}+2x\mathbf{j}\big),\qquad\frac{\partial\mathbf{F}}{\partial y}(x,y)=\big(-\mathbf{i}+(-2y+4y^3)\mathbf{j}\big).$$

Hence

$$\mathbf{DF}(1,0)(x-1,y)=\frac{\partial\mathbf{F}}{\partial x}(1,0)(x-1)+\frac{\partial\mathbf{F}}{\partial y}(1,0)\,y=(3\mathbf{i}+2\mathbf{j})(x-1)-\mathbf{i}\,y$$

so

$$\mathbf{F}(x,y)=\mathbf{F}(1,0)+\mathbf{DF}(1,0)(x-1,y)+\mathbf{R}_2(1,0)(x-1,y)$$
$$=3\mathbf{i}+\mathbf{j}+(3\mathbf{i}+2\mathbf{j})(x-1)-\mathbf{i}\,y+\mathbf{R}(1,0)(x-1,y)$$
$$=(3x-y)\,\mathbf{i}+(2x-1)\mathbf{j}-\mathbf{R}_2(1,0)(x-1,y)$$

In fact, the remainder

$$\mathbf{R}_2(1,0)(x-1,y)=(1+x-y+x^2)\mathbf{i}+(x^2-y^2+y^4)\mathbf{k}-\Big((3x-y)\,\mathbf{i}+(2x-1)\mathbf{j}\Big)$$
$$=(1+x^2-2x)\mathbf{i}+\big(x^2-2x+1+y^2+y^4\big)\mathbf{j}=(x-1)^2\mathbf{i}+\big((x-1)^2+y^2+y^4\big)\mathbf{j}$$

is at least quadratically small in $(x-1)$ and y, i.e. it is bounded by a constant times $(x-1)^2+y^2$, when $(x-1)^2+y^2$ is small.

Def The **Hessian** of a function $f:\mathbf{R}^n\to\mathbf{R}$ is the $n\times n$ matrix

$$Hf=\begin{bmatrix} f_{x_1x_1} & \cdots & f_{x_1x_n}\\ \vdots & \ddots & \vdots\\ f_{x_nx_1} & \cdots & f_{x_nx_n}\end{bmatrix}$$

The second order Taylor formula for a function $f:\mathbf{R}^n\to\mathbf{R}$ can hence be written:

$$f(\mathbf{x})=f(\mathbf{a})+\mathbf{D}f(\mathbf{a})(\mathbf{x}-\mathbf{a})+\frac{1}{2}(\mathbf{x}-\mathbf{a})^THf(\mathbf{a})(\mathbf{x}-\mathbf{a})+R_2(\mathbf{a},\mathbf{x})$$

where T stands for transpose; $(\mathbf{x}-\mathbf{a})^THf(\mathbf{a})(\mathbf{x}-\mathbf{a})=(\mathbf{x}-\mathbf{a})\cdot Hf(\mathbf{a})(\mathbf{x}-\mathbf{a})$.