Higher order derivatives

$f : D \subseteq \mathbb{R}^n \longrightarrow \mathbb{R}^m$. Then

$f' : D \longrightarrow M_{n,m} \cong \mathbb{R}^{nm}$

n inputs, m outputs

$f'' : D \longrightarrow M_{n^2 m} = \mathbb{R}^{n^2 m}$

$f^{(k)} : D \longrightarrow \mathbb{R}^{n^k m}$

Recall: $f' : D \to \mathbb{R}^{nm}$ is differentiable ~~at~~ is defined with the usual defn.

Def f is k times differentiable if $f^{(k-1)}$ is differentiable.

Remark the matrix $f^{(k)} \in M_{n^k, m}$ can be seen to have k inputs of length n and output m

More precisely

$$f^{(k)}(x) = \begin{pmatrix} \frac{\partial^k f_1}{\partial x_1 \dots \partial x_k}(x) \\ \vdots \\ \frac{\partial f_m}{\partial x_1 \dots \partial x_k}(x) \end{pmatrix} \begin{pmatrix} v_1 \\ \vdots \\ v_k \end{pmatrix}$$

Note
This has to be evaluated correctly because of recursive ~~Structure~~

Theorem

$$\frac{\partial^k f}{\partial v_1 \dots \partial v_k}(x) = f^{(k)}(x)$$

$$= \sum_{1 \le j_1 \dots j_k \le n} \frac{\partial^k f}{\partial x_{j_1} \dots \partial x_{j_k}}(x)\, v^1_{j_1} \dots v^k_{j_k}.$$

More precisely $f^{k}(x)(v_1 \dots v_k)$

$$= \frac{\partial}{\partial v_k} f^{k-1}(x)(v_1 \dots v_{k-1})$$

$\boxed{\text{Viewed as a function of } x}$

Quadratic approximation

We consider x_0 and $v = x - x_0$

$$g_v(t) = f(x_0 + tv)$$

$$g_v(1) = g_v(0) + g_v'(0) + \frac{g_v''(1)}{2} + \underbrace{R(v)}_{\text{error term of higher order.}}$$

However

$$g_v'(0) = \frac{\partial f}{\partial v}(x_0) = f'(x_0)\,v = \sum_{j=1}^{k} \frac{\partial f}{\partial x_j}(x_0)\,v_j$$

$$g_v''(0) = \frac{d}{dt}\left(f'(x_0 + tv)\,v\right)\Big|_{t=0}$$

$$= \sum_{l=1}^{k} \sum_{j=1}^{k} \frac{\partial^2 f}{\partial x_j \partial x_l}(x_0)\, v_j v_l$$

Hence **Quadratic Approximation**

$$f(x) \cong f(x_0) + f'(x_0)(x - x_0) + \frac{1}{2} \sum_{l,j=1}^{k} \frac{\partial^2 f}{\partial x_j \partial x_l} (x - x_0)_j (x - x_0)_k$$

$$= f(x_0) + f'(x_0)(x - x_0) + \frac{1}{2} f''(x_0)(x - x_0, x - x_0)$$

This pencast was created with my Livescribe™ smartpen. Click here to see how to create your own.

Theorem Let f be twice differentiable such that

$\cancel{\forall x}$ $f''(x)(v,v) \geq 0$ for all v

Then

$$f(\lambda x + (1-\lambda) y) \leq \lambda f(x) + (1-\lambda) f(y)$$

(f is convex)

Proof Fix x and y and define $v = \cancel{y} - \cancel{x}$,

$$g_v(t) = f(\cancel{x} + t v)$$

Note that

$$f(\lambda x + (1-\lambda) y) = f(y + \lambda(x-y))$$

$$= g_v(\lambda)$$

Using 1D we know that $g_v''(\lambda) \geq 0$ for all λ

implies $g_v(\lambda) = g_v(\lambda 1 + (1-\lambda) 0)$

$$\le \lambda g_v(1) + (1-\lambda) g_v(0)$$

$$= \lambda f(x) + (1-\lambda) f(y)$$

However

$$g_v''(\lambda) = f''(x+\lambda v)(v,v) \ge 0 \quad \text{by assumption} \quad \boxtimes$$

Application

1) By Silvester's theorem a matrix

(a_{jk}) satisfies $\sum h_j a_{jk} h_k \ge 0$

for all h if

all the subdeterminants

a_{11}

$\begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}$

$\begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{32} \end{bmatrix}$

$\vdots$

$$a = \begin{bmatrix} \quad \end{bmatrix}$$

are all positive.

The function $f(x_1 \dots x_n) = \sqrt{1+x_1^2+\dots+x_n^2}$ is convex

Indeed, $\dfrac{\partial f}{\partial x_j} = \dfrac{x_j}{\sqrt{1+\sum x_j^2}}$

$$\frac{\partial^2 f}{\partial x_j \partial x_k} = \begin{cases} -\dfrac{x_j x_k}{(1+\sum x_j^2)^{3/2}} & j \neq k \\ \dfrac{1}{\sqrt{1+\sum x_j^2}} - \dfrac{x_k^2}{(1+\sum x_j^2)^{3/2}} & j = k \end{cases}$$

Therefor $g_{jk} = \left(\dfrac{\partial^2 f}{\partial x_j \partial x_k}\right)\left(\sqrt{1+\sum x_j^2}\right)^{3/2}$

is given by

$$\begin{bmatrix} 1+\sum_{j\neq k} x_k^2 & & -x_j x_k \\ & \ddots & \\ -x_{jk} & & \ddots \end{bmatrix}$$

$n=2$ $\begin{bmatrix} 1+x_2^2 & -x_1 x_2 \\ -x_1 x_2 & 1+x_1^2 \end{bmatrix}$

determinant: $(1+x_1^2)(1+x_2^2) + x_1^2 x_2^2 \geq 0$

The general case also works

$n=3$

$$\begin{matrix} 1+x_2^2+x_3^2 & -x_1x_2 & -x_1x_3 \\ -x_1x_2 & 1+x_1^2+x_3^2 & -x_2x_3 \\ -x_1x_3 & -x_2x_3 & 1+x_1^2+x_2^2 \end{matrix}$$

det

$$= (1+x_2^2+x_3^2)\left[(1+x_1^2+x_3^2)(1+x_1^2+x_2^2) + x_2^2x_3^2\right]$$

$$+ x_1x_2\left(-x_1x_2(1+x_1^2+x_2^2) - x_1x_2x_3^2\right)$$

$$-x_1x_3\left(x_1x_2^2x_3 + x_1x_3(1+x_1^2+x_3^2)\right)$$

At the end there are more + than − terms . . .

<u>Remark</u> Silvester's theorem also shows that if all the minors are strictly positive, then $\sum a_{jk} h_j h_k \geq \delta \sum h_j^2$ for some δ.

Then $g_v''(x_0) > 0$ for every direction, and hence the theorem applies for local max and minima.

Theorem f twice differentiable and, $f'(x_0) = 0$, and

$$\sum \frac{\partial^2 f(x_0)}{\partial x_j \partial x_k} h_j h_k \geq \delta \sum h_j^2$$

Then f has a local minimum

$\boxed{\text{(Pf as in 1D)}}$