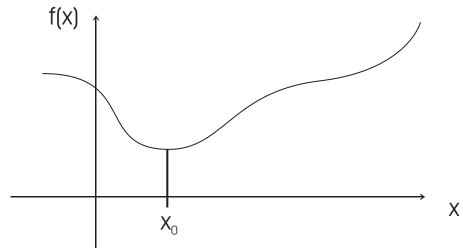
4.7 Minimization problems

Remember 1D: $f(x)$ extreme value at x_0
$f(x) \geq f(x_0)$ for all x close to $x_0 \to x_0$-minimum
(similarly. $f(x) \leq f(x_0) \to$ maximum)
(+ using $f''(x)$, i.e. -curvature information)

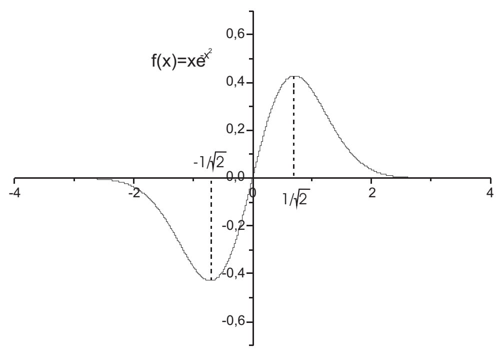


$$x_0\text{-extremum} \Rightarrow f'(x_0) = 0 \wedge \begin{cases} \text{if } f''(x_0) < 0 & \text{maximum} \\ \text{if } f''(x_0) > 0 & \text{minimum} \\ \text{if } f''(x_0) = 0 & \text{no decision,} \\ & \text{(saddle point?)} \end{cases}$$

Example:

$$\begin{aligned} f(x) &= xe^{-x^2} \\ \to f'(x) &= e^{-x^2}(1-2x^2) \to x_0 = \pm\frac{1}{\sqrt{2}} \Leftrightarrow f'(x_0) = 0 \\ \to f''(x) &= e^{-x^2}(2x^2-3)2x \\ \to f''(x_0) &= e^{-\frac{1}{2}}\left(2\frac{1}{2}-3\right)2\frac{\pm 1}{\sqrt{2}} \\ &= -4e^{-\frac{1}{2}}\frac{\pm 1}{\sqrt{2}} = \pm 2\sqrt{2}e^{-\frac{1}{2}} \end{aligned}$$

$$\begin{aligned} f''(+\tfrac{1}{\sqrt{2}}) < 0 &\quad \to \quad \text{max} \\ f''(-\tfrac{1}{\sqrt{2}}) > 0 &\quad \to \quad \text{min} \end{aligned}$$



Example: $f : \mathbb{R}^N \to \mathbb{R}$
$N = 2$ $f(x,y) = x^2 + y^2 - 2x - 4y + 5$
if (x_0, y_0) with $f(x_0, y_0) \leq f(x,y)$ for all (x,y) exist then (x,y) is a minimum:

$$f(x,y) = (x^2 - 2x + 1) + (y^2 - 4y + 4) = (x-1)^2 + (y-2)^2$$

$$f(x,y) \geq 0 \;\to\; x_0 = 1 \wedge \; y_0 = 2 \text{ is a minimum}$$

Derivatives:

$$\begin{aligned} \frac{\partial f}{\partial x} &= 2x - 2 \qquad \frac{\partial f}{\partial x} = 0 \to x_0 = 1 \\ \frac{\partial f}{\partial y} &= 2y - 4 \qquad \frac{\partial f}{\partial y} = 0 \to y_0 = 2 \\ &\to \text{ general feature ? YES!!} \end{aligned}$$

Definition 41 *local minimum $\vec{x}_0$ of $f : \mathbb{R}^N \to \mathbb{R}$ means that $f(\vec{x}) \geq f(\vec{x}_0)$ for all $\vec{x} \in \mathbb{R}^N$ "close to" $\vec{x}_0$.*
local maximum $\vec{x}_0$ of $f : \mathbb{R}^N \to \mathbb{R}$ means that $f(\vec{x}) \leq f(\vec{x}_0)$ for all $\vec{x} \in \mathbb{R}^N$ "close to" $\vec{x}_0$. Calculation of a (local) minimum or maximum in N-dimensions:
If $\vec{x}_0$ is a minimum or maximum then

$$\vec{\nabla} f(\vec{x}_0) = \vec{0}, \; i.e. \; \frac{\partial f(\vec{x}_0)}{\partial x_k} = 0 \text{ for all } k = 1, \ldots, N$$

Note: this is a necessary condition and not always sufficient!! (same as in 1D)
Examples:

(i) Minimum

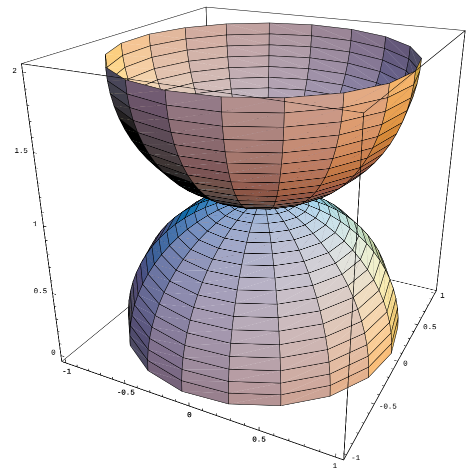
$$f(x,y) = x^2 + y^2 \to \frac{\partial f}{\partial x} = 2x,\; \frac{\partial f}{\partial y} = 2y \quad \to (0,0) \text{ is a minimum since } f(x,y) \geq 0 \text{ for all } (x,y)$$

(ii) Maximum

$$f(x,y) = -x^2 - y^2 \to \frac{\partial f}{\partial x} = -2x,\; \frac{\partial f}{\partial y} = -2x \quad \to (0,0) \text{ is a maximum since } f(x,y) \leq 0 \text{ for all } (x,y)$$

(iii) Saddle point

$$f(x,y) = x^2 - y^2 \quad \to \quad \frac{\partial f}{\partial x} = 2x,\ \frac{\partial f}{\partial y} = -2y$$
$$\to \quad \vec{\nabla} f = \vec{0} \text{ at } \vec{x}_0 = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$



But:

$$\left.\begin{array}{lcl} f(x, y=c) & = & x^2 - c^2 \to \infty \quad \text{for } x \to \infty \\ f(x=c, y) & = & c^2 - y^2 \to -\infty \quad \text{for } y \to \infty \end{array}\right\} \text{ no extreme}$$

$\to$"saddle point" at $(0,0)$ $\Rightarrow$ second derivative?
$\vec{\nabla} f$ is the total derivative of $f : \mathbb{R}^N \to \mathbb{R}$
$\vec{\nabla} f$ is a function $\mathbb{R}^N \to \mathbb{R}^N$ since the gradient is a vector, thus

$$\vec{\nabla} f = \begin{pmatrix} \frac{\partial f}{\partial x_1} \\ \vdots \\ \frac{\partial f}{\partial x_N} \end{pmatrix}$$

$\to$ second (total) derivative is the (total) derivative of the gradient, i.e. it is an $N \times N$ Matrix $\tilde{H}$!

$$\tilde{H}(\vec{x}) = \begin{pmatrix} \frac{\partial^2 f}{\partial x_1^2} & \cdots & \frac{\partial^2 f}{\partial x_1 \partial x_N} \\ \vdots & \ddots & \vdots \\ \frac{\partial^2 f}{\partial x_N \partial x_1} & \cdots & \frac{\partial^2 f}{\partial x_N^2} \end{pmatrix}$$

$\tilde{H}(\vec{x})$ is symmetrical since $\frac{\partial^2 f}{\partial x_j \partial x_k} = \frac{\partial^2 f}{\partial x_k \partial x_j}$
$\tilde{H}(\vec{x})$ is called the second derivative f and has the name "Hesse matrix"