Lecture 2:
Some facts from analysis
Optimality conditions

Optimization T2023

Màster de Fonaments de Ciència de Dades

Universitat de Barcelona

Scalar product

Let $\boldsymbol{x} = (x_1, \ldots, x_n)^T, \boldsymbol{y} = (y_1, \ldots, y_n)^T \in \mathbb{R}^n$.

- Scalar euclidean (dot) product.

$$< \boldsymbol{x}, \boldsymbol{y} >:= \boldsymbol{x} \cdot \boldsymbol{y} = \boldsymbol{x}^T \boldsymbol{y} = x_1 y_1 + \cdots + x_n y_n \in \mathbb{R}.$$

- Euclidean norm.

$$\|\boldsymbol{x}\| = \sqrt{< \boldsymbol{x}, \boldsymbol{y} >} = \sqrt{\boldsymbol{x} \cdot \boldsymbol{x}} = \sqrt{\boldsymbol{x}^T \boldsymbol{x}} = \sqrt{x_1^2 + \cdots + x_n^2}.$$

- Euclidean distance.
  $d(\boldsymbol{x}, \boldsymbol{y}) = \|\boldsymbol{y} - \boldsymbol{x}\| = \sqrt{(y_1 - x_1)^2 + \cdots + (y_n - x_n)^2}.$
- Cosinus of the angle.

$$\cos(\widehat{\boldsymbol{x}, \boldsymbol{y}}) = \frac{\boldsymbol{x} \cdot \boldsymbol{y}}{\|\boldsymbol{x}\|\|\boldsymbol{y}\|}.$$

- Perpendicularity (orthogonality): $\boldsymbol{x} \perp \boldsymbol{y} \quad \Leftrightarrow \quad \boldsymbol{x} \cdot \boldsymbol{y} = 0.$

Cross product

Let $\boldsymbol{x} = (x_1, x_2, x_3), \boldsymbol{y} = (y_1, y_2, y_3) \in \mathbb{R}^3$, we define:

- Cross product.

$$\boldsymbol{x} \times \boldsymbol{y} = \begin{vmatrix} \boldsymbol{i} & \boldsymbol{j} & \boldsymbol{k} \\ x_1 & x_2 & x_3 \\ y_1 & y_2 & y_3 \end{vmatrix} = \begin{pmatrix} x_2 y_3 - x_3 y_2 \\ x_3 y_1 - x_1 y_3 \\ x_1 y_2 - x_2 y_1 \end{pmatrix}.$$

Lemma.

$$\boldsymbol{x} \times \boldsymbol{y} \perp \boldsymbol{x} \quad \text{and} \quad \boldsymbol{x} \times \boldsymbol{y} \perp \boldsymbol{y}.$$

Lines in $\mathbb{R}^2$

The line determined by the point $\boldsymbol{a} = (a_1, a_2)^T$ and the vector $\boldsymbol{v} = (v_1, v_2)^T$ writes as $(t \in \mathbb{R})$

$$\boldsymbol{x} = \boldsymbol{a} + t\boldsymbol{v} \quad \Leftrightarrow \quad \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} a_1 \\ a_2 \end{pmatrix} + t \begin{pmatrix} v_1 \\ v_2 \end{pmatrix}.$$

Equivalently (euclidean coordinates)

$$\frac{x - a_1}{v_1} = \frac{y - a_2}{v_2} \quad \Leftrightarrow \quad Ax + By + C = 0,$$

with $A = v_2,\ B = -v_1,\ C = -a_1 v_2 + a_2 v_1$.

Lines in $\mathbb{R}^3$

- The line determined by the point $\boldsymbol{a} = (a_1, a_2, a_3)^T$ and the vector $\boldsymbol{v} = (v_1, v_2, v_3)^T$ writes as $(t \in \mathbb{R})$

$$\boldsymbol{x} = \boldsymbol{a} + t\boldsymbol{v} \quad \Leftrightarrow \quad \begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} a_1 \\ a_2 \\ a_3 \end{pmatrix} + t \begin{pmatrix} v_1 \\ v_2 \\ v_3 \end{pmatrix}.$$

Equivalently (euclidean coordinates)

$$\frac{x - a_1}{v_1} = \frac{y - a_2}{v_2} = \frac{z - a_3}{v_3}.$$

Planes in $\mathbb{R}^3$

The plane determined by the point $\boldsymbol{a} = (a_1, a_2, a_3)^T$ and the vectors $\boldsymbol{u} = (u_1, u_2, u_3)^T$ and $\boldsymbol{v} = (v_1, v_2, v_3)^T$ writes as $(t,\ s \in \mathbb{R})$

$$\boldsymbol{x} = \boldsymbol{a} + t\boldsymbol{u} + s\boldsymbol{v} \quad \Leftrightarrow \quad \begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} a_1 \\ a_2 \\ a_3 \end{pmatrix} + t \begin{pmatrix} u_1 \\ u_2 \\ u_3 \end{pmatrix} + s \begin{pmatrix} v_1 \\ v_2 \\ v_3 \end{pmatrix}.$$

The above equation of the plane can also be written as

$$\begin{vmatrix} x - a_1 & y - a_2 & z - a_3 \\ u_1 & u_2 & u_3 \\ v_1 & v_2 & v_3 \end{vmatrix} = 0$$

or as

$$Ax + By + Cz + D = 0,$$

with $(A, B, C)^T = \boldsymbol{u} \times \boldsymbol{v}$.

Real-valued functions: continuity

Consider the real valued function

$$f : \mathbb{R}^n \longrightarrow \mathbb{R}.$$

- The domain $D := D(f)$ of f is the set of points $\boldsymbol{x} \in \mathbb{R}^n$ where f is defined.
- The graph of f, as the subset of $\mathbb{R}^{n+1}$, is defined as

$$\mathrm{graf}(f) := \left\{(\boldsymbol{x}, z) \in \mathbb{R}^{n+1} : \begin{array}{l} \boldsymbol{x} = (x_1, \ldots, x_n)^T \in D \subset \mathbb{R}^n, \\ z = f(\boldsymbol{x}) \in \mathbb{R} \end{array} \right\}.$$

- The level set of f (of level $c \in \mathbb{R}$) is given by

$$L_c = \{\boldsymbol{x} \in D \,:\, f(\boldsymbol{x}) =\} \subset \mathbb{R}^n.$$

Real-valued functions: continuity (cont.)

We say that f is continuous at a point $\boldsymbol{a} \in \mathcal{C}$ if and only if

$$\lim_{\boldsymbol{x} \to \boldsymbol{a}} f(\boldsymbol{x}) = f(\boldsymbol{a}).$$

(whenever the limit has meaning)

- The elementary functions of one (or several) variable

  $$x^n, \quad e^x, \quad \log x, \quad \sin x, \quad \cos x$$

  are continuous in their domain.
- Addition, subtraction, product, division (except at the points where the denominator vanishes) and composition of continuous functions are also continuous functions.

Real-valued functions: continuity (cont.)

Theorem (Bolzano). Let $f : [a, b] \to \mathbb{R}$ be a continuous function. Assume that $f(a)f(b) < 0$. Then there exists $c \in (a, b)$ such that $f(c) = 0$.

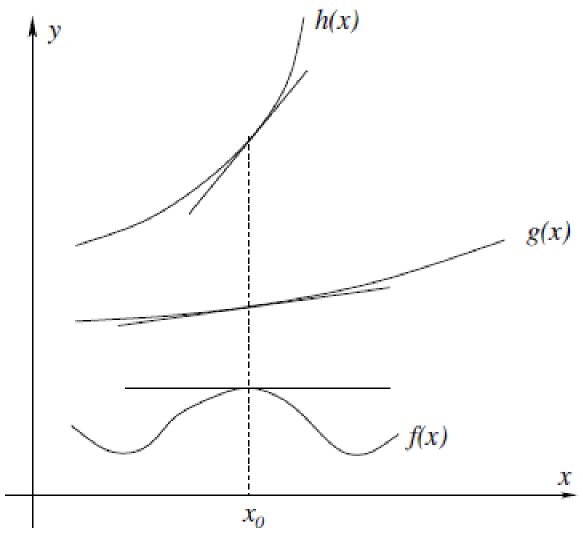
Theorem (Weierstrass). Let

$$f : K \subset \mathbb{R}^n \longrightarrow \mathbb{R},$$

be continuous function such that K is compact (closed and bounded), then f is bounded (there exist M such that $|f(x)| < M$ for all $x \in K$) and f attaints its maximum and minimum values on K.

Derivative in dimension $n = 1$

The derivative of a function $y = f(x)$ is a measure of the (infinitesimal) rate at which the value y of the function changes with respect to the change of the variable x.



- Geometrically. The limit of the secant lines is the tangent line and the derivative is the slope of the tangent line.
- Analitically. The derivative of the function f at the point a is

$$f'(a) = \lim_{h \to 0} \frac{f(a+h) - f(a)}{h}.$$

Derivative in dimension $n = 1$

Easily, one can check that the derivative satisfies the property that

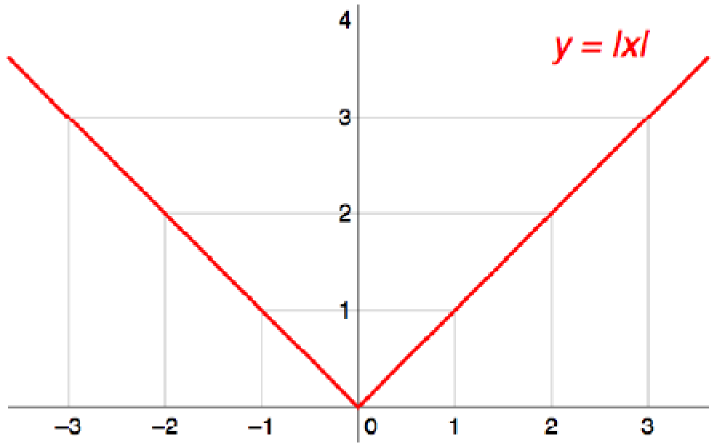
$$\lim_{h \to 0} \frac{f(a+h) - f(a) - f'(a) \cdot h}{h} = 0,$$

which has the intuitive interpretation that the tangent line to f at a gives the best linear approximation to f near a (h small).

$$f(a+h) \approx f(a) + f'(a)h,$$

Definition. We say that f is differentiable at the point $a \in D(f)$ if there exists the derivative of f at the point a (the limit exists).

Continuity and differentiability in dimension $n = 1$



- Theorem. If f is differentiable at the point $a \in D(f)$ then f is continuous at the point a.
- The converse is not necessarily true.

We say that $f : (a, b) \to \mathbb{R}$ is differentiable (in (a, b)) if it is differentiable at every point.

Differentiability in $\mathbb{R}^n$, $n \geq 1$: partial derivatives

Let f be a real valued function that depends on n variables

$$\begin{aligned} f: \quad \mathbb{R}^n &\longrightarrow \mathbb{R} \\ \boldsymbol{x} &\longrightarrow f(\boldsymbol{x}) = f(x_1, \ldots, x_n), \end{aligned}$$

Let $\boldsymbol{a} = (a_1, \ldots, a_n)^T$ be an interior point of $D(f)$. We want to extend the notion of differentiability introduced above for $n = 1$.

Partial derivative(s). The partial derivative of $f(\boldsymbol{x})$ in the direction x_j at the point $\boldsymbol{a}$ is defined to be:

$$f_{x_j}(\boldsymbol{a}) := \frac{\partial f}{\partial x_j}(\boldsymbol{a}) = \lim_{h \to 0} \frac{f(a_1, \ldots, a_j + h, \ldots, a_n) - f(a_1, \ldots, a_j, \ldots, a_n)}{h}.$$

Differentiability in $\mathbb{R}^n$, $n \geq 1$: gradient vector

The gradient vector. Let f and $\boldsymbol{a}$ as before. The vector formed by the partial derivatives of f at the point a (assuming it exists) is known as the gradient vector

$$\nabla f(a) = \left(\frac{\partial f}{\partial x_1}(\boldsymbol{a}), \ldots, \frac{\partial f}{\partial x_n}(\boldsymbol{a}) \right)^T$$

Differentiability in $\mathbb{R}^n$. Let f and $\boldsymbol{a}$ as before. We say that f is differentiable at $\boldsymbol{a}$ if

$$\lim_{||\boldsymbol{h}|| \to 0} \frac{\left| f(\boldsymbol{a} + \boldsymbol{h}) - f(\boldsymbol{a}) - (\nabla f(a))^T \boldsymbol{h} \right|}{||h||} = 0, \quad h \in \mathbb{R}^n.$$

One can show that if partial derivatives of f are continuous functions at $\boldsymbol{a}$, then f is (continuous) differentiable ($\mathcal{C}^1$ map).

Differentiability in $\mathbb{R}^n$, $n \geq 1$: directional derivative

The partial derivatives of f measure the variation in f in the axis directions. But in many occasions we need to measure the variation of f in any direction, represented by a vector $v \in \mathbb{R}^n$. One can show that w.l.o.g. we may assume $||v|| = 1$ (unitary vector).

Choose a unitary vector $\boldsymbol{v} = (v_1, \ldots, v_n)^T$. The directional derivative of f in the direction of $\boldsymbol{v}$ at the point $\boldsymbol{a}$ is defined by

$$D_{\boldsymbol{v}} f(\boldsymbol{a}) = \lim_{h \to 0} \frac{f(\boldsymbol{a} + h\boldsymbol{v}) - f(\boldsymbol{a})}{h}.$$

Differentiability in $\mathbb{R}^n$, $n \geq 1$: directional derivative (cont.)

Theorem. Let $\alpha : (a, b) \longrightarrow \mathbb{R}^n$ be a differentiable curve, $\alpha(t) = (\alpha_1(t), \ldots, \alpha_n(t))^T$. Let f be a real valued differentiable function as above. Then

$$f(\alpha(t)) = f(\alpha_1(t), \ldots, \alpha_n(t)),$$

and

$$\frac{d}{dt} f(\alpha(t)) = \frac{\partial f}{\partial x_1}(\alpha(t))\alpha_1'(t) + \cdots + \frac{\partial f}{\partial x_n}(\alpha(t))\alpha_n'(t).$$

Corollary. In the above notation (assume f is differentiable) we have

$$D_{\mathbf{v}} f(\mathbf{x}) = \left.\frac{d}{dt}\right|_{t=0} f(\mathbf{x} + t\mathbf{v}) = \sum_{j=1}^{n} \frac{\partial f(\mathbf{x})}{\partial x_j} v_j =< (\nabla f(\mathbf{a}))^T, v > .$$

The gradient vector

Theorem. Let $f : D \subset \mathbb{R}^n \longrightarrow \mathbb{R}$ be a differentiable function at $\boldsymbol{a} \in D$, and $\boldsymbol{u} \in \mathbb{R}^n$ is an unitary vector. The following statements hold.

(a)
$$D_{\boldsymbol{u}} f(\boldsymbol{a}) = (\nabla f(\boldsymbol{a})) \cdot \boldsymbol{u} = \|\nabla f(\boldsymbol{a})\| \cos\theta,$$
where θ is the angle between $\boldsymbol{u}$ and $\nabla f(\boldsymbol{a})$.

(b) The gradient vector $\nabla f(\boldsymbol{a})$ gives the maximum direction variation of f at the point $\boldsymbol{a}$.

(c) The gradient vector at the point $\boldsymbol{a} \in D$ is orthogonal to the level curve passing through $\boldsymbol{a}$.

Proof. Statements (a) and (b) are direct. Let $\boldsymbol{r}(t) = (x_1(t), x_2(t), \ldots, x_n(t))^T$ be the level curve passing through $\boldsymbol{a} \in D$. Then

$$0 = \frac{d}{dt} f(\boldsymbol{r}(t)) = \cdots = < \nabla f(\boldsymbol{r}(t))^T, \boldsymbol{r}'(t) > .$$

Gradient vector: linear approximation

- If $n = 1$ the linear approximation of the function f at a point $\boldsymbol{a} \in D \subset R$ is defined by the linear function

$$L(x) = f(x_0) + f'(x_0)(x - x_0).$$

- In dimension n we have

$$L(\boldsymbol{x}) = f(\boldsymbol{x}_0) + (\nabla f(\boldsymbol{x}_0))^T(\boldsymbol{x} - \boldsymbol{x}_0).$$

Gradient vector: linear approximation

- If $n = 1$ the linear approximation of the function f at a point $\boldsymbol{a} \in D \subset R$ is defined by the linear function

$$L(x) = f(x_0) + f'(x_0)(x - x_0).$$

- In dimension n we have

$$L(\boldsymbol{x}) = f(\boldsymbol{x}_0) + (\nabla f(\boldsymbol{x}_0))^T(\boldsymbol{x} - \boldsymbol{x}_0).$$

Example: Let $f(\boldsymbol{x}) = f(x, y, z) = ze^x\sqrt{y}$. Estimate the value of $f(0.01, 24.8, 1.02)$.

We take $\boldsymbol{a} = (0, 25, 1)^T$ and we use the linear approximation of f given by

$$L(\boldsymbol{x}) = f(\boldsymbol{a}) + (\nabla f(\boldsymbol{a}))^T(\boldsymbol{x} - \boldsymbol{a}) = 5 + (5,\ 1/10,\ 5)\begin{pmatrix} x - 0 \\ y - 25 \\ z - 1 \end{pmatrix}$$

$$= 5 + 5x + \frac{1}{10}(y - 25) + 5(z - 1)$$

Finally

$$L(0.01, 24.8, 1.02) = 5.13 \approx f(0.01, 24.8, 1.02) = 5.1306.$$

Real-valued functions: higher-order derivatives

Let f be a real valued function that depends on n variables

$$\begin{array}{rccl} f: & D \subset \mathbb{R}^n & \longrightarrow & \mathbb{R} \\ & \boldsymbol{x} & \longrightarrow & f(\boldsymbol{x}^T) = f(x_1, \ldots, x_n), \end{array}$$

Let $\boldsymbol{a} = (a_1, \ldots, a_n)$ be an interior point of $D(f)$. Assume that not only f but all its partial derivatives are also differentiable functions at a point $\boldsymbol{a} \in D$.

We consider the derivatives of order two (second order derivatives) as follows

$$\frac{\partial^2 f}{\partial x_i \partial x_j}(\boldsymbol{a}) := \frac{\partial}{\partial x_i}\left(\frac{\partial f}{\partial x_j}\right)(\boldsymbol{a}).$$

Real-valued functions: the Hessian

Let $f : D \subset \mathbb{R}^n \longrightarrow \mathbb{R}$ be a real valued function. Assume D is open and that f admits up to second order derivatives. Then we define the Hessian matrix at the point $\boldsymbol{a} \in D$ as follows

$$\nabla^2(f)(a) \quad := \begin{pmatrix} \frac{\partial^2 f}{\partial x_1^2}(\boldsymbol{a}) & \cdots & \frac{\partial^2 f}{\partial x_n \partial x_1}(\boldsymbol{a}) \\ \vdots & \ddots & \vdots \\ \frac{\partial^2 f}{\partial x_1 \partial x_n}(\boldsymbol{a}) & \cdots & \frac{\partial^2 f}{\partial x_n^2}(\boldsymbol{a}) \end{pmatrix}$$

Theorem (Schwartz's Lemma). If f admits up to second order derivatives in D and those functions are continuous in D (f is $\mathcal{C}^2$), then the Hessian matrix is symmetric. Indeed,

$$\frac{\partial^2 f}{\partial x_i \partial x_j}(\boldsymbol{a}) = \frac{\partial^2 f}{\partial x_j \partial x_i}(\boldsymbol{a}), \quad a \in D.$$

Taylor's expansion for f : linear part

Assume all needed regularity of f in the next slides. When we say that

$$L(\boldsymbol{x}) = f(\boldsymbol{a}) + (\nabla f(\boldsymbol{a}))^T (\boldsymbol{x} - \boldsymbol{a})$$

is the linear approximation of the function f at a point $\boldsymbol{a}$ we mean (roughly speaking) that

$$f(\boldsymbol{x}) = f(\boldsymbol{a}) + (\nabla f(\boldsymbol{a}))^T (\boldsymbol{x} - \boldsymbol{a}) + \mathcal{O}\left(|\boldsymbol{x} - \boldsymbol{a}|^2\right).$$

This is known as Taylor's expansion of order one.

Taylor's expansion for f : quadratic part

The second order approximation of f at the point $\boldsymbol{a} \in D$ is

$$f(\boldsymbol{x}) = f(\boldsymbol{a}) + (\nabla f(\boldsymbol{a}))^T (\boldsymbol{x} - \boldsymbol{a}) + \frac{1}{2} \nabla^2(f)(\boldsymbol{a})(\boldsymbol{x} - \boldsymbol{a})^2 + \mathcal{O}\left(|\boldsymbol{x} - \boldsymbol{a}|^3\right)$$

where the value of $\nabla^2(f)(\boldsymbol{a})(\boldsymbol{x} - \boldsymbol{a})^2 \in \mathbb{R}$ is given by

$$(x_1 - a_1, \ldots, x_n - a_n) \begin{pmatrix} \frac{\partial^2 f(\boldsymbol{a})}{\partial x_1^2} & \cdots & \frac{\partial^2 f(\boldsymbol{a})}{\partial x_n \partial x_1} \\ \vdots & \ddots & \vdots \\ \frac{\partial^2 f(\boldsymbol{a})}{\partial x_1 \partial x_n} & \cdots & \frac{\partial^2 f(\boldsymbol{a})}{\partial x_n^2} \end{pmatrix} \begin{pmatrix} x_1 - a_1 \\ \vdots \\ x_n - a_n \end{pmatrix}.$$

We say that

$$Q(\boldsymbol{x}) = f(\boldsymbol{a}) + (\nabla f(\boldsymbol{a}))^T (\boldsymbol{x} - \boldsymbol{a}) + \tfrac{1}{2} \nabla^2(f)(\boldsymbol{a})(\boldsymbol{x} - \boldsymbol{a})^2$$

is the quadratic approximation of the function f at a point $\boldsymbol{a}$.

Quadratic functions

As it will be clear during the course it is crucial for getting the optimality conditions to determine whether the expression $\nabla^2(f)(\boldsymbol{a})(\boldsymbol{x}-\boldsymbol{a})^2$ is positive or negative for $\boldsymbol{x} \in \mathbb{R}^n$ (or at least for $\boldsymbol{x}$ near $\boldsymbol{a}$). Remember that $\nabla^2(f)(\boldsymbol{a})$ is a symmetric matrix. The key notion is the following.

Definition. Let Q any symmetric $n \times n$ matrix.

1. Q is positive semidefinite (PSD) if and only if $\boldsymbol{x}^T Q\boldsymbol{x} \geq 0$ for all $\boldsymbol{x} \in \mathbb{R}^n$.
2. Q is positive definite (PD) if and only if $\boldsymbol{x}^T Q\boldsymbol{x} > 0$ for all $\boldsymbol{x} \in \mathbb{R}^n$, $\boldsymbol{x} \neq 0$.

Theorem. Let Q any symmetric $n \times n$ matrix. Then all aigenvalues of Q are real. Moreover Q is PD if and only if all eigenvalues of Q are positive.

Local and global minima

Let $D \subset \mathbb{R}^n$ an open set. Let $f : D \longrightarrow \mathbb{R}$ be a real valued function. Denote by $\mathrm{B}(\boldsymbol{x}^\star, \varepsilon) := \{x \in \mathbb{R}^n \mid \|\boldsymbol{x} - \boldsymbol{x}^*\| \leq \varepsilon\}$.

1. We say that $\boldsymbol{x}^\star \in D$ is a local minimum of f if there exists $\epsilon > 0$ such that $\mathrm{B}(\boldsymbol{x}^\star, \varepsilon) \subset D$ and $f(\boldsymbol{x}^*) \leq f(\boldsymbol{x})$ for all $\boldsymbol{x} \in \mathrm{B}(\boldsymbol{x}^\star, \varepsilon)$. If $f(\boldsymbol{x}^*) < f(\boldsymbol{x})$ for all $\boldsymbol{x} \in \mathrm{B}(\boldsymbol{x}^\star, \varepsilon) \setminus \{\boldsymbol{x}^\star\}$ we say that $\boldsymbol{x}^\star$ is a strict local minimum of f.
2. We say that $\boldsymbol{x}^\star \in D$ is a global minimum of f if $f(\boldsymbol{x}^*) \leq f(\boldsymbol{x})$ for all $\boldsymbol{x} \in D$. If $f(\boldsymbol{x}^*) < f(\boldsymbol{x})$ for all $\boldsymbol{x} \in D \setminus \{\boldsymbol{x}^\star\}$ we say that $\boldsymbol{x}^\star \in D$ is a strict global minimum of f.
3. Similar definitions correspond to (local and global) maxima of f instead of minima of f. Finally we write extrema of f when we refer to max or min, indiscriminately.

Local and global minima

Theorem : Let $f : [a, b] \to \mathbb{R}$ be a continuous function. Let $c \in (a, b)$. If for some $\delta > 0$, f is decreasing on $(c - \delta, c)$ and increasing on $(c, c + \delta)$, then f has a local minimum at c.

Proof: Let x_1 and x such that $c - \delta < x_1 < x < c$. Then

$$f(x_1) \geq \lim_{x \to c^-} f(x) = f(c).$$

Similarly, if $c < x < x_2 < c + \delta$, then

$$f(x_2) \geq \lim_{x \to c^+} f(x) = f(c).$$

Exercise. Prove that the condition of the theorem is not necessary.

Necessary condition for minima

Theorem. Let $D \subset \mathbb{R}^n$ open and let $f : D \to \mathbb{R}$ be a $\mathcal{C}^1$ function (continuously differentiable). Assume $\boldsymbol{x}^\star \in D$ is a local minima of f. Then

$$\nabla f(\boldsymbol{x}^*) = 0.$$

Proof. We argue by contradiction. By hypothesis, if $\lambda > 0$ is small enough, for all $\boldsymbol{v} \in \mathbb{R}^n$, we have that

$$f(\boldsymbol{x}^*) \leq f(\boldsymbol{x}^* + \lambda \boldsymbol{v}). \tag{1}$$

Fix $\boldsymbol{v}$ (w.l.o.g. $||v|| = 1$). Define $F_{\boldsymbol{v}}(\lambda) := F(\lambda) = f(\boldsymbol{x}^* + \lambda \boldsymbol{v})$. Then, (1) writes as

$$F(0) \leq F(\lambda), \quad \forall\, |\lambda| < \delta.$$

for some $\delta > 0$. From the Mean Value Theorem, we have

$$F(\lambda) = F(0) + F'(\theta\lambda)\lambda,$$

for some $\theta \in [0, 1]$.

Proof (continue).

If $F'(0) > 0$, then, since F' is a continuous function, we have that, if $|\lambda|$ is small enough, $F'(\theta\lambda) > 0$ for all $\theta \in [0,1]$. Hence, we can find a $\lambda < 0$ small enough such that

$$F(\lambda) = F(0) + F'(\theta\lambda)\lambda < F(0),$$

a contradiction. Arguing similarly with $F'(0) < 0$ we conclude $F'(0) = 0$, or equivalently,

$$F'(0) = (\nabla f(\boldsymbol{x}^\star))^T \boldsymbol{v} = 0.$$

Since $\boldsymbol{v}$ is an arbitrary unitary vector, we must have:

$$\nabla f(\boldsymbol{x}^\star) = 0.$$

Remark. Similar arguments show that if $\boldsymbol{x}^\star \in D$ is a local maxima of a $\mathcal{C}^1$ map f , then

$$\nabla f(\boldsymbol{x}^*) = 0.$$

Sufficient conditions for minima

Theorem. Let $D \subset \mathbb{R}^n$ open and let $f : D \to \mathbb{R}$ be a $\mathcal{C}^2$ map (twice continuously differentiable). Assume

$$\nabla f(\boldsymbol{x}^*) = 0, \quad \mathbf{z}^T \, \nabla^2(f)(\mathbf{x}^*) \, \boldsymbol{z} > 0, \quad \forall \boldsymbol{z} \neq 0.$$

Then f has a strict local minimum at $\boldsymbol{x}^*$.

Exercise. Prove the Theorem above (Use the second order Taylor expansion of f around $\boldsymbol{x}^*$).

Exercise. Prove that the converse of Theorem above is not true. Consider the family of one-dimensional real valued functions $f_p(x) = x^p, \; p \geq 1$.

Necessary and sufficient conditions for minima

Theorem A. Let $D \subset \mathbb{R}^n$ open and let $f : D \to \mathbb{R}$ be a $\mathcal{C}^2$ map (twice continuously differentiable). The following statements hold

(a) (Necessary conditions) Assume $\boldsymbol{x}^\star \in D$ is a local minima of f. Then

$$\nabla f(\boldsymbol{x}^*) = 0 \quad \text{and} \quad \mathbf{z}^T\, \nabla^2(f)(\mathbf{x}^*)\, \boldsymbol{z} \geq 0, \qquad \forall \boldsymbol{z} \in \mathbb{R}^n.$$

(b) (Sufficient conditions) Assume $\nabla f(\boldsymbol{x}^*) = 0$. Assume also that there exists $\varepsilon > 0$ such that for all $\boldsymbol{x} \in \boldsymbol{B}(\boldsymbol{x}^\star, \varepsilon)$ we have

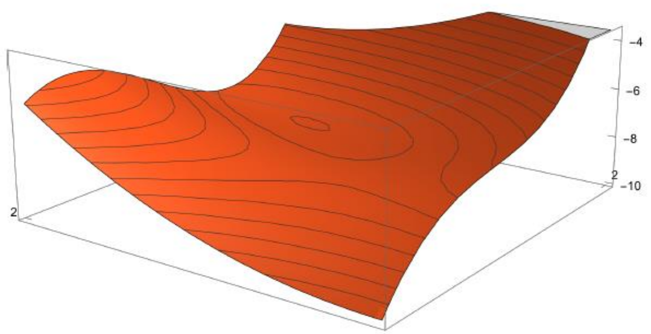
$$\mathbf{z}^T\, \nabla^2(f)(\mathbf{x})\, \boldsymbol{z} \geq 0, \qquad \forall \boldsymbol{z} \in \mathbb{R}^n.$$

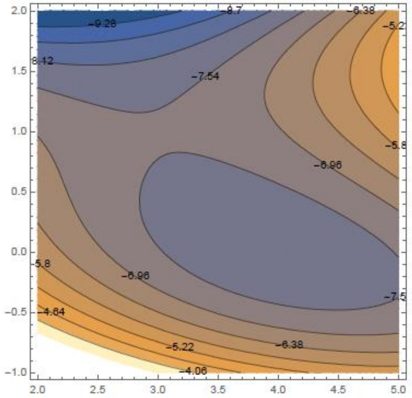
Then, $\boldsymbol{x}^\star \in D$ is a local minima of f.

An example

Problem. Find the (local) extrema for the polynomial function

$$f(x,y) = \frac{1}{2}x^2 + xy + 2y^2 - 4x - 4y - y^3.$$





Computations. We have that

$$\nabla f(x,y) = (x+y-4,\ x+4y-4-3y^2)^T \quad \text{and} \quad \nabla^2(f)(x,y) = \begin{pmatrix} 1 & 1 \\ 1 & 4-6y \end{pmatrix}.$$

Easily

$$\nabla f(x,y) = 0 \quad \Longleftrightarrow \quad \mathbf{x_1} = (4,0)^T \text{ and } \mathbf{x_2} = (3,1)^T,$$

and

$$\nabla^2 f(\mathbf{x_1}) = \begin{pmatrix} 1 & 1 \\ 1 & 4 \end{pmatrix}, \quad \nabla^2 f(\mathbf{x_2}) = \begin{pmatrix} 1 & 1 \\ 1 & -2 \end{pmatrix}.$$

$\nabla^2 f(\mathbf{x_1})$ is positive definite , $\nabla^2 f(\mathbf{x_2})$ is indefinite.

Conclusion. The only extrema is the local minimum is $\mathbf{x_1} = (4,0)^T$.

Problem (exercise). Find the extrema for the polynomial function

$$f(x,y) = \left(4 - \frac{21}{10}x^2 + \frac{1}{3}x^4\right)x^2 + xy + 4y^2(-1+y^2)$$