Master-MIRI
Topics on Optimization and Machine Learning
(TOML)

José M. Barceló Ordinas
Departament d'Arquitectura de Computadors
(UPC)

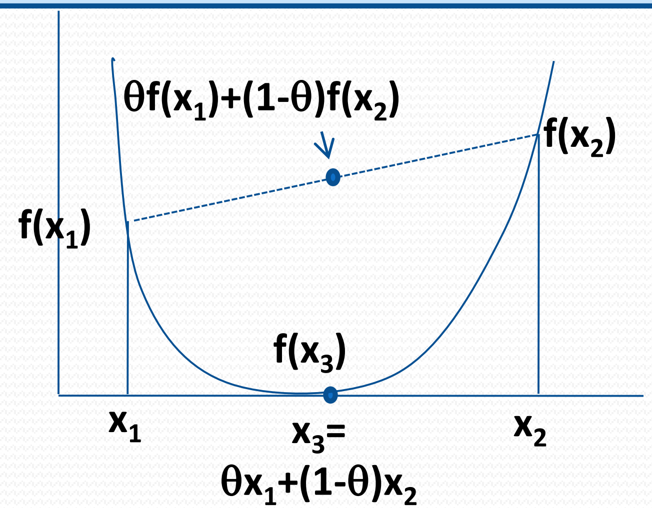
Domains of a function and Convex Functions

A function f: A$\longrightarrow$ B is a **mapping** between sets A and B. The **domain** of a function is the "input" parameters of the function, it is to say, all $x \in \mathrm{dom}\, f \subset A$ if $f(x) \subset B$ exists.

A **function** f: $R^n \longrightarrow R$ is **convex** if

i. For all $x \in \mathrm{dom}\, f \subset R^n$, then dom f is a convex set
ii. For $0 \leq \theta \leq 1$, we have $f(\theta x_1 + (1-\theta) x_2) \leq \theta f(x_1) + (1-\theta) f(x_2)$

- Strict convexity → change "≤" for "<"
- A function "**f**" is concave if "**–f**" is convex
- Affine functions, f(x)=Ax+b (and therefore also linear functions, f(x()=Ax), hold equality in condition ii) and thus are both convex and concave



Linear functions and affine functions

In **analytic geometry**, **a linear function** is a polynomial: e.g. in one dimension $f(x) = ax+b$ or in more dimensions $f(x_1, \ldots x_n) = a_1 x_1 + \ldots + a_n x_n$ is a hyperplane.

In **linear algebra, a linear function (or linear map)** is a mapping between 2 vector spaces that preserves addition and scalar multiplication: $f(\alpha x+\beta y) = \alpha f(x) + \beta f(y)$, where x, y are vectors

An **affine function** is the composition of a translation and a linear map. For example, if A is a matrix, $f(x) = Ax$ is a linear function (linear map) and affine and $f(x) = Ax+b$ is an affine function (but not linear).

- **Reminder**
  - Let f: $R^n \longrightarrow R$ a function

The function f is **differentiable** if dom f is open and there exists the partial derivative (is a vector) at each point $x \in \text{dom } f \subset R^n$

$$\nabla f(x)=(\delta f(x)/\delta x_1, \delta f(x)/\delta x_2, \ldots, \delta f(x)/\delta x_n)$$

The function is twice differentiable if the dom f is open and the **Hessian** $\nabla^2 f$ (is a matrix) exists at each point $x \in \text{dom } f \subset R^n$

$$\nabla^2 f(x)_{ij} = \delta f^2(x)/\delta x_1 \delta x_2$$

A matrix A is **positive semi-definite** iif $\forall x$, $x^T A x \geq 0$ (**positive definite** if $x^T A x > 0$). Ways of checking whether a matrix is positive semi-definite/definite is:

- All eigenvalues of A are ≥ 0 (positive definite→ all are >0)
- All leading principal minors have positive or equal to cero determinants (positive definite→ all are >0)

A matrix A is **negative semi-definite** iif $\forall x$, $x^T A x \leq 0$ (**negative definite** if $x^T A x < 0$). Ways of checking whether a matrix is negative semi-definite/definite is:

- All eigenvalues of A are ≤ 0 (negative definite→ all are <0)
- All leading **odd** principal minors have **negative** or equal to cero determinants and all **even** principal minors have **positive** or equal to cero determinants (negative definite→ odd are <0 and even are >0)

- **Reminder**

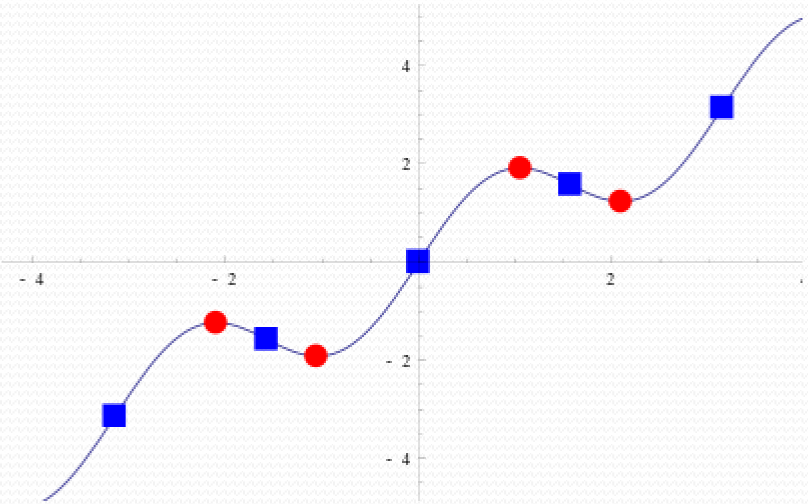
Critical point of a function of a real variable is any value in the domain where either the function is not differentiable or its derivative is 0.
If the derivative is zero, the point is called a **stationary point** of the function. Then a stationary point is a critical point but not all critical points are stationary (e.g. there is no derivative).

An **inflection point** (is a stationary point) is a point on a curve at which the curvature or concavity changes sign from plus to minus or from minus to plus, i.e., a point on a curve at which the second derivative changes sign and the first derivative is 0.
Local maxima and minima of a function can occur only at its critical points. But, not every stationary point is a maximum or a minimum of the function, e.g. not at inflection points.

- **Reminder**

If the second derivative is positive is a minimum (stationary point) and if it is negative it is a maximum (stationary point).

If the second derivative is zero, the nature of the stationary point must be determined by way of other means, often by noting a sign change around that point provided the function values exist around that point.



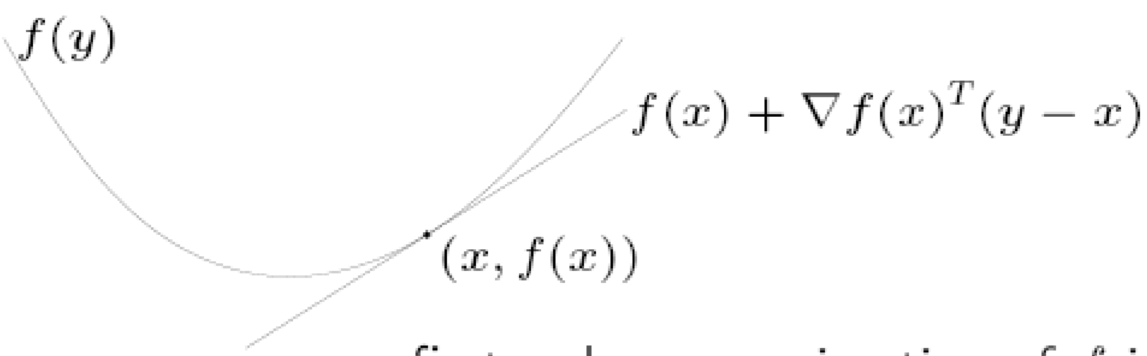
Blue squares: inflection points
Red dots: maximum or minimum

First-order conditions

Let $f: R^n \longrightarrow R$ be differentiable, it is to say ∇f exists in the whole domain of f. Then f is convex iif

i. **dom f** is a convex set,
ii. $\mathbf{f(y) \geq f(x)+\nabla f(x)^T(y-x)}$ (first order Taylor approximation of f near x) for all x, $y \in$ dom f

From **local information** about a convex function, a **global information** is obtained.



first-order approximation of f is global underestimator

Second-order conditions

Let $f: R^n \longrightarrow R$ be twice differentiable, it is to say the Hessian $\nabla^2 f$ exists in the whole domain of f. Then f is convex iif

i. **dom f** is a convex set,
ii. $\mathbf{H=\nabla^2 f(x) \geq 0}$, the Hessian matrix is positive semi-definite.

Be careful, condition i) is necessary: $f(x)=1/x^2$ in dom $f=\{x\in R,\ x\neq 0\}$ has $f''(x)>0$ for all $x\in$ dom f, but is not convex.

Examples:

- All **affine and linear functions** are convex, and concave functions,
- **Quadratic functions:** $f(x)=½x^TPx+q^Tx+r$ are convex for all $P\geq 0$ (positive semi-definite) matrices and $x\in R^n$,
- **Exponential functions:** e^{ax} is convex on R and any $a\in R$,

- **More examples:**
  - **Powers of absolute value** $|x|^a$ with $a \geq 1$ are convex on R,
  - **Logarithms** log x is concave on R++,
  - **Negative entropy** x logx is convex on R+ (0log0=0),
  - Any **norm** $||\cdot||_p$ is convex on R^n,
  - **Max function**, $f(x)=\max\{x_1,x_{2,\ldots,}\ x_n\}$ is convex on R^n,
  - **Quadratic over linear function** $f(x,y)=x^2/y$ is convex on RxR++,
  - **Log-sum-exp**, $f(x)=\log(\ e^{x1}+e^{x2}+\ldots+e^{xn})$ is convex on R^n,
  - **Geometric mean** $f(x)=(\Pi^n{}_{i=1}\ x_i)^{1/n}$ is concave on R^n ++,
  - **Log-determinant**, f(x)=log (det A) is concave on S^n ++ where S^n ++ is the set of symmetric positive definite nxn matrices.

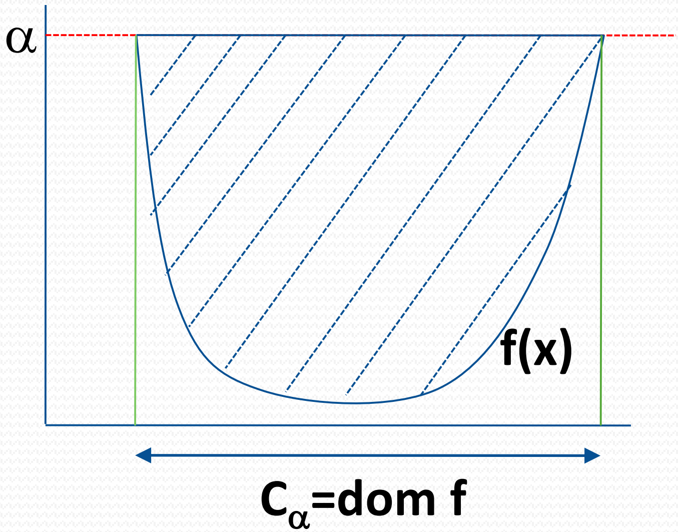
- ## Sub-level sets

A **α-sublevel** set of a function f: $R^n \longrightarrow R$ is defined as

$$C_\alpha=\{x\in \text{dom } f \mid f(x)\leq\alpha\}$$

Where we define **α-level** sets when equality. **Sublevel sets** of a convex function **are convex** for any α (converse is false: convexity of a sublevel does not imply convexity in the function).

If f is concave, **α-superlevel** sets are defined as $C_\alpha=\{x\in \text{dom } f \mid f(x)\geq\alpha\}$ and are convex. d



Proof:

If $x_1,x_2\in C_\alpha$, then $f(x_1)\leq\alpha$, $f(x_2)\leq\alpha$

Then:

$$f(\theta x_1+(1-\theta)x_2) \leq \theta f(x_1)+(1-\theta)f(x_2) \leq$$
$$\leq \theta\alpha+(1-\theta)\alpha \leq \alpha \quad \text{for } 0\leq\theta\leq1$$

Then $\theta x_1+(1-\theta)x_2\in C_\alpha$

- **α-level set of a quadratic function:**
  - $f(x) = \frac{1}{2} x^T P x$ with P positive definite matrix. Then:

$$C^{=}_{\alpha} = \{x \in \mathrm{dom}\, f \mid \tfrac{1}{2} x^T P x = \alpha\}$$

Is an ellipsoid with center 0.

  - Proof:

$$f(x) = \tfrac{1}{2} x^T P x + q^T x = \tfrac{1}{2}(x + P^{-1} q)^T P (x + P^{-1} q) - \tfrac{1}{2} q^T P^{-1} q$$

Then, the level set $C^{=}_{\alpha} = \{x \in \mathrm{dom}\, f \mid f(x) = \alpha\}$ forms an ellipsoid of center $x_0 = -P^{-1} q$

Remember the equation of an ellipsoid:

$$E = \{ x \mid \tfrac{1}{2}(x - x_0)^T P (x - x_0) = \alpha\}$$

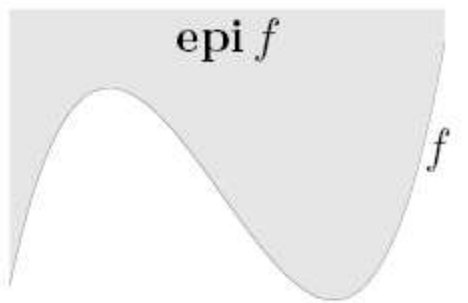
- ## Epigraphs/Hypographs

The graph of a function $f: R^n \longrightarrow R$ is defined as

$$\{(x,f(x)) \mid x \in \text{dom } f\}$$

The **epigraph** of a function $f: R^n \longrightarrow R$ is defined as

$$\text{epi } f = \{(x,t) \mid x \in \text{dom } f \text{ and } f(x) \leq t\} \subset R^{n+1}$$



The **hypograph** of a function $f: R^n \longrightarrow R$ is defined as

$$\{(x,t) \mid x \in \text{dom } f \text{ and } f(x) \geq t\} \subset R^{n+1}$$

A **function is convex** iif its epigraph is a **convex set**
A **function is concave** iif its hypograph is a **convex set**

The epigraph definition gives another tool to test whether a function is convex or not

- **Jensen inequality**

The inequality $f(\theta x_1+(1-\theta)x_2) \leq \theta f(x_1)+(1-\theta)f(x_2)$ also is called **Jensen inequality** and can be extended to k points, if f convex and $x_1,x_2,\ldots,x_k \in \mathrm{dom}\ f$ and $\theta_1+\theta_2+\ldots+\theta_k=1$, then

$$f(\theta_1 x_1+\theta_2 x_2+\ldots+ \theta_k x_k) \leq \theta_1 f(x_1)+ \theta_1 f(x_1)+ \ldots+\theta_k f(x_k)$$

and as in the case of convex sets, this inequality extends to infinite sums, integrals and expectations, e.g.:

- $f(\int_S p(x)\, x\, dx) \leq \int_S f(x)\, p(x)\, dx$ if the integral exists
- $f(E(x)) \leq E(f(x))$ where $E(\cdot)$ is the expectation of r.v. x

Conjugate Function

The function $f^*: R^n \to R$ is called the **conjugate function**, with f^* defined as:

$$f^*(y) = \sup_x (\langle x,y \rangle - f(x)) = \sup_x (y^T x - f(x))$$

The conjugate function is convex

- It can be interpreted as the negative of the y-intercept of the tangent line to the graph of f that has slope y. In other words, we look for the largest affine function below f, it is to say, the one with largest intercept.
- It can also be interpreted as an encoding of the convex hull of the function's epigraph in terms of its supporting hyperplanes.

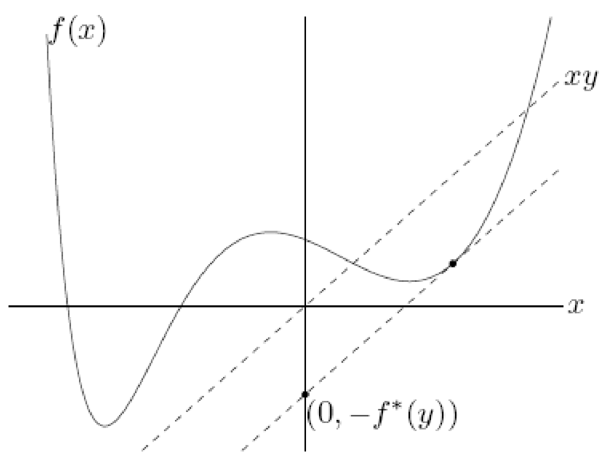


Figure 3.8 A function $f : \mathbf{R} \to \mathbf{R}$, and a value $y \in \mathbf{R}$. The conjugate function $f^*(y)$ is the maximum gap between the linear function yx and $f(x)$, as shown by the dashed line in the figure. If f is differentiable, this occurs at a point x where $f'(x) = y$.

It can also be interpreted as an encoding of the convex hull of the function's epigraph in terms of its supporting hyperplanes.

$$f^*(y) = \sup_x (\langle x,y\rangle - f(x)) = \sup_x (y^T x - f(x))$$

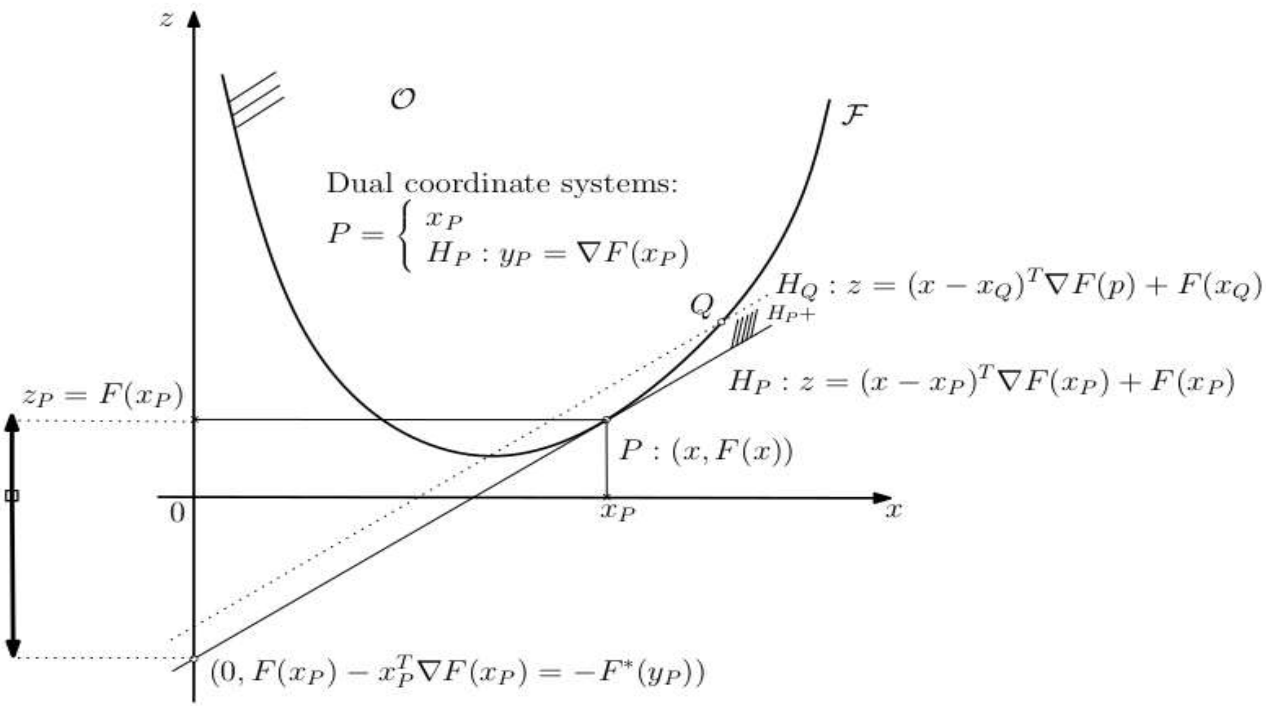


Figure 1: Illustration the Legendre transformation of a strictly convex function: A point P on the boundary of $\mathcal{O}$ can either be parameterized by using the x-coordinate system, or by using the dual slope $y = \nabla F(x)$ coordinate system. For a point $P \in \partial O$ with x-coordinate x_P, and tangent parameter $y_P = \nabla F(x_P)$, the Legendre conjugate $F^*(y)$ reads as the intersection of the hyperplane H_P with the the z-axis. The object $\mathcal{O}$ is either interpreted as the convex hull of the points, or dually as the intersection of the supporting half-spaces.

Conjugate Function

The function $f^*: R^n \to R$ is called the **conjugate function**, with f^* defined as:

$$f^*(y) = \sup_x (\langle x,y\rangle - f(x)) = \sup_x (y^T x - f(x))$$

The conjugate function is convex
It can be interpreted as the negative of the y-intercept of the tangent line to the graph of f that has slope y. In other words, we look for the largest affine function below f, it is to say, the one with largest intercept.

- If f is convex and differentiable → **Legendre transform**: the x value that attains the maximum satisfies that $y=\nabla_x f(x)$ and then:

$$f^*(y) = \sup_x (y^T x - f(x)) = (x^T \nabla_x f(x) - f(x)) = f^*(\nabla_x f(x))$$

- $f^{**}=f$ iif f is convex and closed (i.e. epi f is closed)

- **Conjugate Function**
  - **Examples**:
    - **Affine function**: $f(x)=ax+b \rightarrow f^*(y)=-b$ at $y=a$
    - **Exponential**: $f(x)=e^x \rightarrow f^*(y)=y\log(y)-y$ with dom $f^*=R+$
    - **Negative entropy:** $f(x)=x\log(x) \rightarrow f^*(y)=e^{y-1}$ with dom $f^*=R$
    - **Indicator function**: $f_S(x)=0$ if $x\in S \rightarrow f^*(y)=\sup_x (y^T x)$
    - **Log-sum-exp**: $f(x)=\log(\Sigma_{i=1..m}\exp(x_i)) \rightarrow f^*(y)=\Sigma_{i=1..m} y_i\log(y_i)$ with $1^T y=1$ and $y\geq 0$

- **Operations that preserve convexity**

  - **Non-negative weighted sums:** $f=w_1f_1+\ldots+w_mf_m$ is convex if f_i is convex and $w_i \geq 0$ for $i=1,\ldots,m$
  - **Non-negative weighted integrals:** $g(x)=\int_A w(y)f(x,y)dy$ is convex if $f(x,y)$ convex in x and $w(y)\geq 0$ for each $y \in A$
  - **Composition with an affine mapping:** $f:R^n \longrightarrow R$, $A \in R^{nxm}$, $b \in R^n$, and let be $g:R^n \longrightarrow R$ such as $g(x)=f(Ax+b)$ with dom $g=\{x \mid Ax+b \in \text{dom } f\}$. Then, if f is convex (concave), so is g.
  - **Pointwise maximum:** $g(x)=\max\{f_1(x),\ldots,f_m(x)\}$ is convex if f_i is convex for $i=1,\ldots,m$
  - **Pointwise supremum:** $g(x)=\sup_{y \in A}\{f(x,y)\}$ is convex if $f(x,y)$ is convex in x for each $y \in A$. The domain of g is, dom $g=\{x \mid (x,y) \in \text{dom } f \text{ for all } y \in A, \sup_{y \in A}\{f(x,y)\}<\infty\}$.

Composition

- Let be $h:R^k \longrightarrow R$ and $g:R^n \longrightarrow R^k$ functions, and let us consider composition $f=h^\circ g=h(g(x))$: $R^n \longrightarrow R^k \longrightarrow R$, with dom $f=\{x \in \text{dom } g \mid g(x) \in \text{dom } h\}$.

- Let us consider the cases, k=1 and n=1, $h:R \longrightarrow R$, $g:R \longrightarrow R$ and remember than $f'(x)=h'(g(x))\ g'(x)$ and that $f''(x)=h'(g(x))\ (g'(x))^2 + h'(g(x))\ g''(x)$. In order to be convex, $f''(x) \geq 0$

f is convex if h is convex and non-decreasing and g is convex
f is convex if h is convex and non-increasing and g is concave
f is concave if h is concave and non-decreasing and g is concave
f is concave if h is concave and non-increasing and g is convex

- Similar conditions for n>1, but with considering the extended-value function of h which assigns values to $\pm\infty$

Examples of Composition

- If g is convex → $f(x)=e^{g(x)}$ is convex
- If g is concave and positive → **log(g(x))** is concave
- If g is concave and positive → **1/g(x)** is convex
- If g is convex and non-negative and $p \geq 1$ → $g(x)^p$ is convex
- If g is convex → **-log(-(g(x))** is convex on $\{x|g(x)<0\}$

Vector Composition

- Let us now consider $h:R^k \longrightarrow R$ and $g_i:R^n \longrightarrow R$ functions, and let us consider composition $f=h(g_1(x),\ldots,g_k(x)):R^n \longrightarrow R$. Considering n=1, we have $f''(x)=g'(x)^T \nabla h(g(x))\ g'(x) + \nabla h(g(x))^T\ g''(x)$, then f(x) is convex if $f''(x)\geq 0$ and

f is convex if h is convex and non-decreasing in each argument and g_i are convex
f is convex if h is convex and non-increasing in each argument and g_i are concave
f is concave if h is concave and non-decreasing in each argument and g_i are concave
f is concave if h is concave and non-increasing in each argument and g_i are convex

Examples of vector composition

- $h(z)=\log(\Sigma_{i=1,..k}\, e^{z_i})$ is convex an non-decreasing at each argument, then $\log(\Sigma_{i=1,..k}\, e^{g_i(x)})$ is convex whatever $g_i(x)$ is
- If $g_i(x)$ are convex and non-negative, then $(\Sigma_{i=1,..k}\, g_i(x)^p)^{1/p}$ is convex

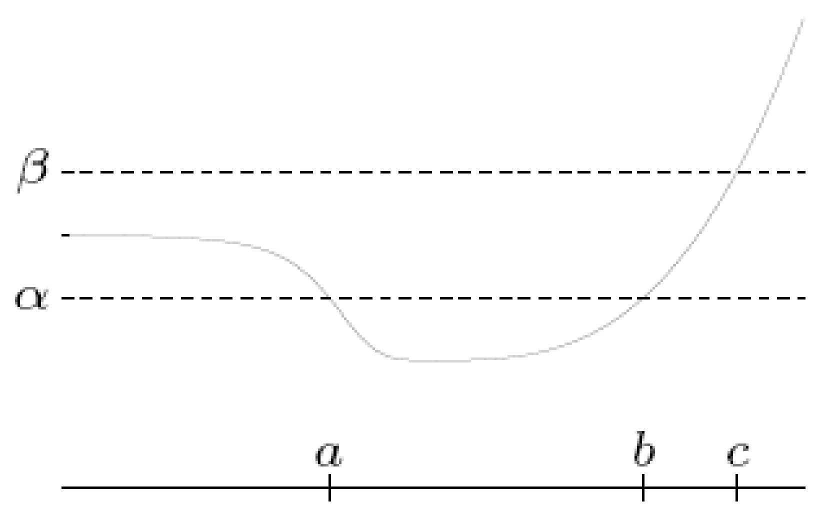
- **Quasi-convex functions**

A function $f:R^n \longrightarrow R$ is called **quasi-convex** if its domain and all its sublevel sets $S_\alpha=\{x\in \text{dom } f \mid f(x)\leq\alpha\}$ for $\alpha\in R$ are convex.

Figure: $S_\alpha=[a,b]$ is convex, $S_\beta=(-\infty,c]$ is convex

A function $f:R^n \longrightarrow R$ is called **quasi-concave** if $-f$ is quasi-convex, e.g. superlevel sets $S_\alpha=\{x\in \text{dom } f \mid f(x)\geq\alpha\}$ for $\alpha\in R$ are convex

A function that is quasi-convex and quasi-concave is called quasi-linear, $S_\alpha=\{x\in \text{dom } f \mid f(x)=\alpha\}$



Characterization of quasi-convex functions

A function $f:R^n \longrightarrow R$ is **quasi-convex** if its domain f is convex and for any $x_1,x_2 \in$ dom f, we have

$$f(\theta x_1+(1-\theta)x_2) \leq \max\{f(x_1),f(x_2)\}$$

The value of the function on a segment does not exceed the value at the endpoints.

This is called **Jensen's inequality for quasi-convex functions**

A function $f:R^n \longrightarrow R$ is **quasi-concave** if its domain f is convex and for any $x_1,x_2 \in$ dom f, we have

$$f(\theta x_1+(1-\theta)x_2) \geq \min\{f(x_1),f(x_2)\}$$

- **Examples of quasi-convex functions**

  - **Logarithm**: log x on R++ is quasi-linear
  - **Ceiling function**: $\text{ceil}(x)=\inf\{z \in Z \mid z \geq x\}$ is quasi-linear
  - **Linear fractional function:** $f(x) = (a^T x+b)/(c^T x+d)$ is quasi-linear in dom $f=\{x \mid c^T x+d>0\}$
  - $f(x_1,x_2)=x_1 x_2$ is quasi-concave in R++
  - **Distance ratio function**: $f(x)=||x-a||_2/||x-b||_2$, then f is quasi-convex on the halfspace $\{x \mid ||x-a||_2 \leq ||x-b||_2\}$