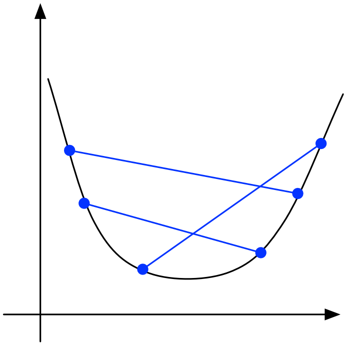
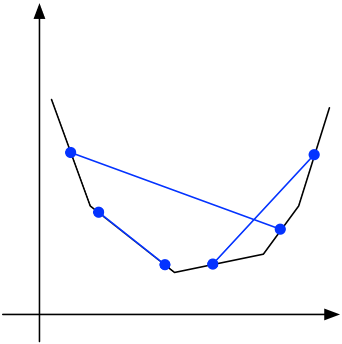
Convex functions

Nonlinear optimization
Instituto Superior Técnico and Carnegie Mellon University PhD course

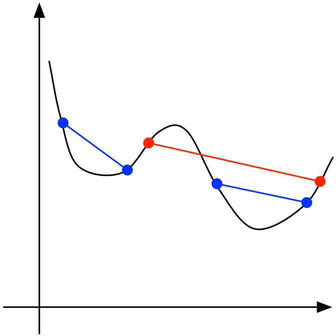
João Xavier
TA: Hung Tuan

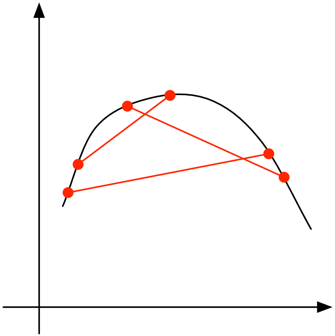
Convex functions





Nonconvex functions





Definition of convex function in a vector space

Definition. A function $f : V \to \mathbf{R} \cup \{+\infty\}$ in a vector space V is convex if $f \not\equiv +\infty$ and

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

for all $x, y \in S$ and $\alpha \in [0,1]$. $(0(+\infty) := 0)$

Definition. The domain of f is

$$\text{dom}\, f = \{x \in V \,:\, f(x) < +\infty\}.$$

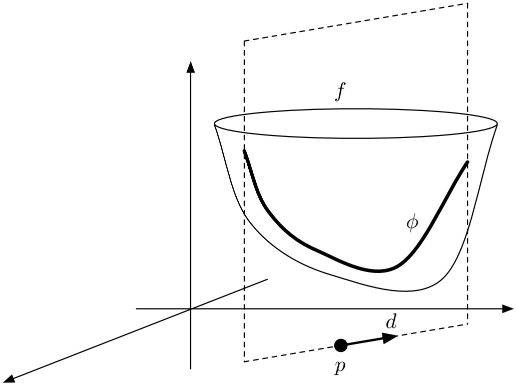
If f is convex, its domain is a convex set

Convexity is a 1D property

Proposition. f is convex if and only $\phi : \mathbf{R} \to \mathbf{R} \cup \{+\infty\}$,

$$\phi(t) = f(p + td),$$

is convex for any p and d in V.



How do we recognize convex functions?

List of simple ones

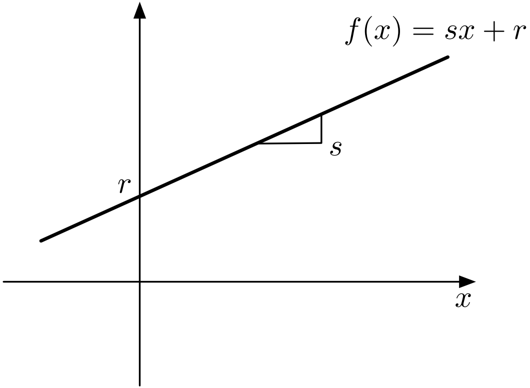
$+$

Apply convexity-preserving operations

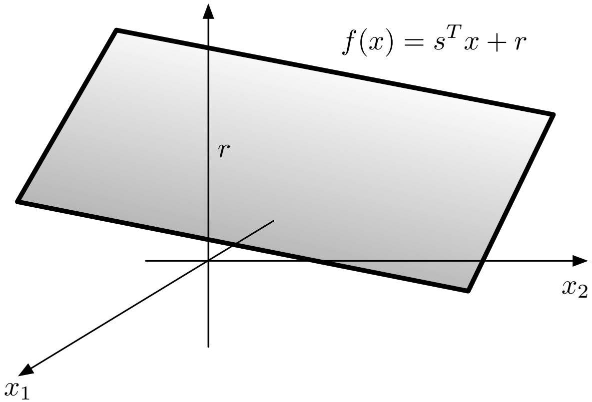
Simple convex functions

- affine
- norms
- indicators

Affine function in $\mathbf{R}$



Affine function in $\mathbf{R}^2$



Affine function

Definition. An affine function $f : V \to \mathbf{R}$ is a map of the form

$$f(v) = l(v) + r$$

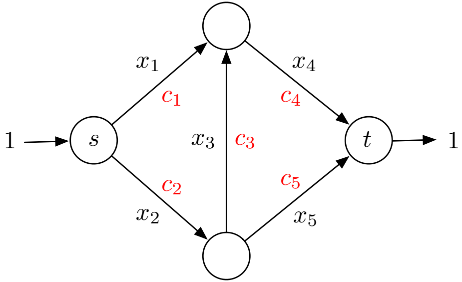
for some linear function $l : V \to \mathbf{R}$ and some $r \in \mathbf{R}$.

Examples:

- $f : \mathbf{R}^n \to \mathbf{R}$, $f(x) = s^T x + r$ $(s \in \mathbf{R}^n, r \in \mathbf{R})$
- $f : \mathbf{R}^{n\times m} \to \mathbf{R}$, $f(X) = \mathrm{tr}\left(S^T X\right) + r$ $(S \in \mathbf{R}^{n\times m}, r \in \mathbf{R})$
- $f : \mathbf{S}^n \to \mathbf{R}$, $f(X) = \mathrm{tr}\left(SX\right) + r$ $(S \in \mathbf{S}^n, r \in \mathbf{R})$

Theorem. An affine function is convex.

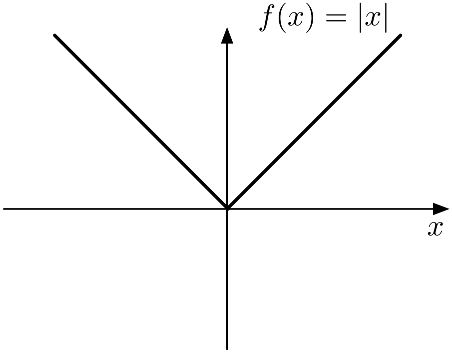
Example: network flow



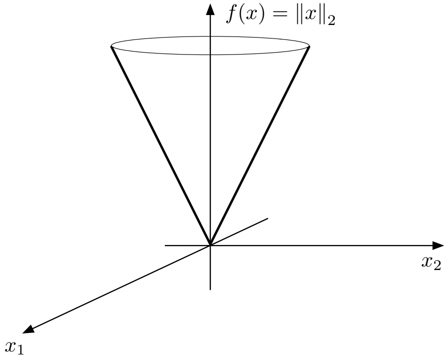
Formulation that minimizes cost:

$$\begin{aligned} \underset{x_1,x_2,x_3,x_4,x_5}{\text{minimize}} \quad & c_1x_1 + c_2x_2 + c_3x_3 + c_4x_4 + c_5x_5 \\ \text{subject to} \quad & 1 = x_1 + x_2 \\ & x_1 + x_3 = x_4 \\ & x_2 = x_3 + x_5 \\ & x_4 + x_5 = 1 \\ & x_1, x_2, x_3, x_4, x_5 \geq 0 \end{aligned}$$

Norm in $\mathbf{R}$



Norm in $\mathbf{R}^2$



Theorem. A norm is a convex function.

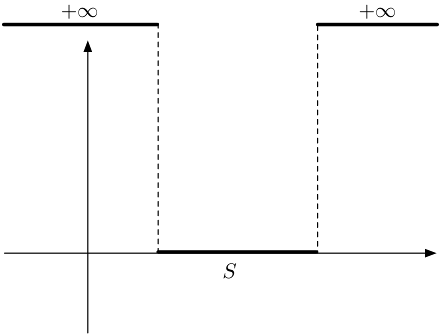
Examples:

- $f : \mathbf{R}^n \to \mathbf{R}$, $f(x) = \|x\|_2$
- $f : \mathbf{R}^n \to \mathbf{R}$, $f(x) = \|x\|_\infty$
- $f : \mathbf{S}^n \to \mathbf{R}$, $f(X) = \|X\|_\mathsf{F}$

Indicators

Definition. The indicator of a set $S \subset V$ is the function

$$i_S : V \to \mathbf{R} \cup \{+\infty\}, \qquad i_S(x) = \begin{cases} 0, & \text{if } x \in S \\ +\infty, & \text{otherwise.} \end{cases}$$



Theorem. The indicator of a convex set is a convex function.

Indicators allow to pass constraints to the objective:

$$\begin{aligned} \underset{x}{\text{minimize}} \quad & f(x) \\ \text{subject to} \quad & x \in A \\ & x \in B \end{aligned}$$

is equivalent to

$$\begin{aligned} \underset{x}{\text{minimize}} \quad & f(x) + i_B(x) \\ \text{subject to} \quad & x \in A \end{aligned}$$

is equivalent to

$$\underset{x}{\text{minimize}} \quad f(x) + i_A(x) + i_B(x)$$

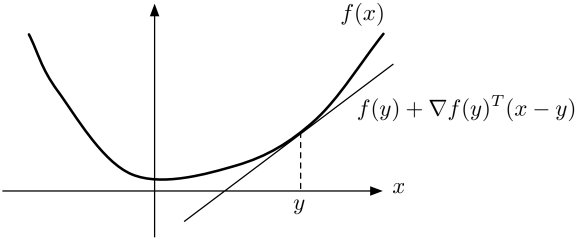
is equivalent to

$$\underset{x}{\text{minimize}} \quad f(x) + i_{A \cap B}(x)$$

Convexity through differentiability

Theorem (1st order criterion). Let S be an open convex subset of $\mathbf{R}^n$ and $f : S \to \mathbf{R}$ be a differentiable function. Then,

$$f \text{ is convex} \quad \Leftrightarrow \quad f(x) \geq f(y) + \nabla f(y)^T (x - y) \text{ for all } x, y \in S.$$



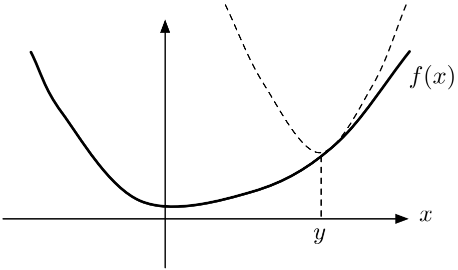
Mostly useful in the $\Rightarrow$ direction:

- offers affine lower bounds to convex functions
- produces interesting inequalities

Convexity through differentiability

Theorem (2nd order criterion). Let S be an open convex subset of $\mathbf{R}^n$ and $f : S \to \mathbf{R}$ be a twice-differentiable function. Then,

$$f \text{ is convex} \quad \Leftrightarrow \quad \nabla^2 f(y) \succeq 0 \text{ for all } y \in S.$$



- Mostly useful in the $\Leftarrow$ direction: proves f is convex
- Commonly, the 2nd order Taylor expansion

$$f(y) + \nabla f(y)^T(x - y) + \frac{1}{2}(y - x)^T \nabla^2 f(y)(y - x)$$

is **not** an upper bound on $f(x)$

Examples

- $f : \mathbf{R}_{++} \to \mathbf{R}$, $f(x) = \frac{1}{x}$
- $f : \mathbf{R}_{++} \to \mathbf{R}$, $f(x) = -\log(x)$
- $f : \mathbf{R}_{+} \to \mathbf{R}$, $f(x) = x\log(x)$ with $0\log(0) := 0$

- $f : \mathbf{R}^n \to \mathbf{R}$, $f(x) = x^T A x + b^T x + c$, with A symmetric, is convex iff $A \succeq 0$
- $f : \mathbf{R}^n \to \mathbf{R}$, $f(x_1, \ldots, x_n) = \log\left(e^{x_1} + \cdots + e^{x_n}\right)$
- $f : \mathbf{R}^n \times \mathbf{R}_{++} \to \mathbf{R}$, $f(x, y) = \frac{x^T x}{y}$

Theorem. For any $A \in \mathbf{S}^n_{++}$ and $B \in \mathbf{S}^n$ there exists $S \in \mathbf{R}^{n \times n}$ such that

$$A = SS^T \quad \text{and} \quad B = S\Lambda S^T$$

where $\Lambda \in \mathbf{R}^{n \times n}$ is diagonal, with the eigenvalues of $A^{-1/2}BA^{-1/2}$.

Further examples of convex functions:

- $f : \mathbf{S}^n_{++} \to \mathbf{R}$, $f(X) = \mathsf{tr}\left(X^{-1}\right)$
- $f : \mathbf{S}^n_{++} \to \mathbf{R}$, $f(X) = -\log\det(X)$

Operations that preserve convexity

- conic combination
- composition with affine map
- pointwise supremum

Conic combination preserves convexity

Theorem. Let $f_i : V \to \mathbf{R} \cup \{+\infty\}$ be convex functions and $\alpha_i \geq 0$ for $i = 1, \ldots, n$. If $\bigcap_{i=1}^n \operatorname{dom} f_i \neq \emptyset$, then

$$f = \alpha_1 f_1 + \cdots + \alpha_n f_n$$

is convex.

Example: basis pursuit with denoising

$$\underset{x}{\text{minimize}} \quad \underbrace{\|Ax - b\|_2^2 + \rho \|x\|_1}_{f(x)}$$

for given $A \in \mathbf{R}^{m \times n}$, $b \in \mathbf{R}^m$, and $\rho > 0$

Composition with affine map preserves convexity

Theorem. Let $A : V \to W$ be an affine map, $f : W \to \mathbf{R} \to \{+\infty\}$ be convex, and $A(V) \cap \operatorname{dom} f \neq \emptyset$. Then

$$f \circ A : V \to \mathbf{R} \cup \{+\infty\}$$

is convex.

Example: logistic regression

- $(x_k, y_k) \in \mathbf{R}^n \times \{0, 1\}$, $k = 1, \ldots, K$, is the labeled training set
- label is generated randomly from feature vector

$$\mathbf{P}(Y = y \mid X = x, s, r) = \frac{e^{y(s^T x + r)}}{1 + e^{s^T x + r}}$$

- what should be the classifier parameters $(s, r) \in \mathbf{R}^n \times \mathbf{R}$?

$$\underset{s,r}{\text{maximize}} \sum_{k=1}^{K} \log \mathbf{P}\left(Y = y_k \mid X = x_k, s, r\right)$$

Pointwise supremum preserves convexity

Theorem. Let $f_i : V \to \mathbf{R} \cup \{+\infty\}$, $i \in \mathcal{I}$, be a family a convex functions and suppose that $f = \sup_{i \in \mathcal{I}} f_i \not\equiv +\infty$. Then, $f : V \to \mathbf{R} \cup \{+\infty\}$ is convex.

Example: fire-station placement

- $p_1, \ldots, p_K$ are the locations of villages
- what should be the position x of the fire-station?

$$\underset{x}{\text{minimize}} \quad \underbrace{\max\{\|x - p_1\|, \cdots, \|x - p_K\|\}}_{f(x)}$$

Example: maximum eigenvalue function

$$\lambda_{\max} : \mathbf{S}^n \to \mathbf{R} \qquad X \mapsto \lambda_{\max}(X)$$

- does not have a closed-form expression
- variational characterization from linear algebra

$$\lambda_{\max}(X) = \sup\{q^T X q \, : \, \|q\| = 1\}$$

offers the representation

$$\lambda_{\max} = \sup\{f_q(X) \, : \, q \in \mathcal{Q}\}$$

with $f_q : \mathbf{S}^n \to \mathbf{R}$, $f_q(X) = q^T X q$ and $\mathcal{Q} = \{q \in \mathbf{R}^n \, : \, \|q\| = 1\}$

- f_q is a convex function (linear function of X)
- we conclude $\lambda_{\max}$ is a convex function