T5: Convex optimization problems

Pablo Arias Martínez - ENS Paris-Saclay, UPF
pablo.arias@upf.edu

October 21, 2021

Optimization and inference techniques for Computer Vision

Previously on...

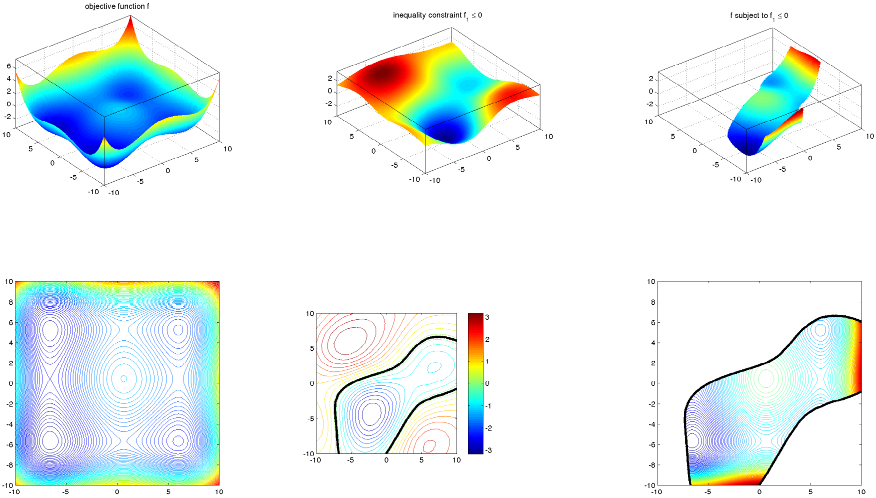
Optimizers:

- In general, Adam allows faster convergence, but it has been reported that SGD generalizes better
- Adabelief
- Centralized gradients

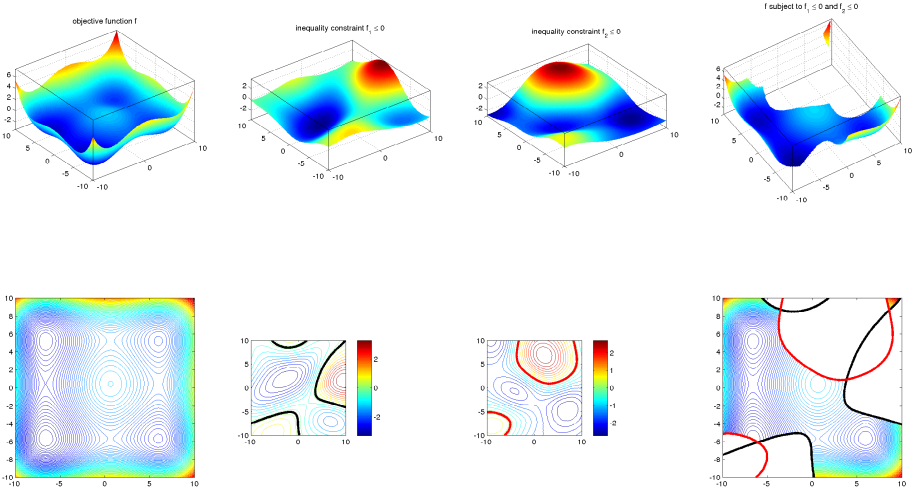
LR scheduling:

- There are other forms of schedulling the LR.
- Cosine annealing
- LR cycling

Constrained optimization

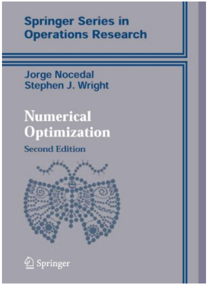


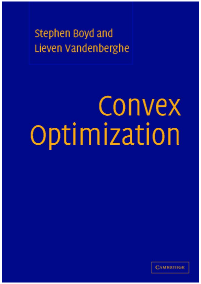
Constrained optimization



Bibliography

- Nocedal, J., Wright, S.J., "Numerical Optimization", Springer.
- Boyd, S., Vandenberghe, L., "Convex Optimization", Cambridge University Press.
  http://www.stanford.edu/~boyd/cvxbook/
- Stanford course on Convex Optimization II
  https://web.stanford.edu/class/ee364b/lectures.html





Constrained optimization problem

Let $C \subset \mathbb{R}^n$ and $f_0 : \mathbb{R}^n \to \mathbb{R}$.

$$\min_{\mathbf{x} \in C} f_0(\mathbf{x}) \quad \longleftrightarrow \quad \begin{array}{ll} \text{minimize} & f_0(\mathbf{x}) \\ \text{subject to} & \mathbf{x} \in C \end{array}$$

Some definitions

- f_0 objective or cost function
- $C \cap \mathrm{dom}(f_0)$ is the feasible set
- $\mathbf{x} \in C \cap \mathrm{dom}(f_0)$ is a feasible point
- if $C \cap \mathrm{dom}(f_0) = \emptyset$ the problem is infeasible
- $p^* = \min_{\mathbf{x} \in C} f_0(\mathbf{x})$ optimal value (minimum)
- $\mathbf{x}$ is optimal if $f(\mathbf{x}) = p^*$

Constrained optimization problem (explicit constraints)

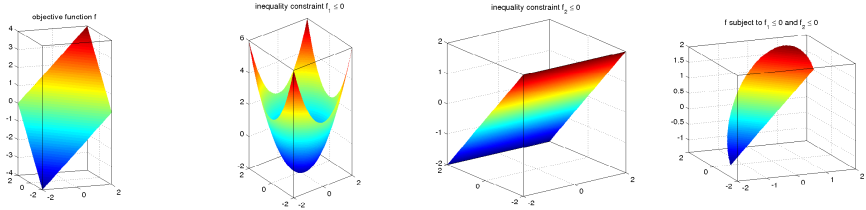
Let $f_0, f_i : \mathbb{R}^n \to \mathbb{R},\ i = 1, \dots, m$.

$$\begin{array}{ll} \text{minimize} & f_0(\mathbf{x}) \\ \text{subject to} & f_i(\mathbf{x}) \leq 0, \quad i = 1, \dots, m. \end{array}$$

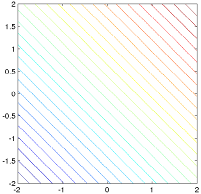
Some definitions

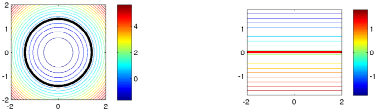
- f_0 objective or cost function
- $C \cap \mathrm{dom}(f_0)$ is the feasible set
- $\mathbf{x} \in C \cap \mathrm{dom}(f_0)$ is a feasible point
- if $C \cap \mathrm{dom}(f_0) = \emptyset$ the problem is infeasible
- $p^* = \min_{\mathbf{x} \in C} f_0(\mathbf{x})$ optimal value (minimum)
- $\mathbf{x}$ is optimal if $f(\mathbf{x}) = p^*$

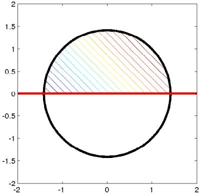
Example of constrained optimization



$$
\begin{array}{ll}
\text{minimize} & x_1 + x_2 \\
\text{subject to} & x_1^2 + x_2^2 - 2 \leq 0 \\
 & -x_2 \leq 0.
\end{array}
$$







Does our problem have a solution?
(Existence)

Does our problem have an unique solution?
(Uniqueness)

How do we know if a point x is a solution?
(Optimality conditions)

Is it possible to find find the solution?
(Convexity)

Can we still find solutions for non-differentiable problems?
(Non-smooth Optimization)

Existence of minimizers: examples of functions with no minimers

When do the solutions of an optimization problem in $\mathbb{R}^n$ exist?

$$\min_{x \in C} f_0(\mathbf{x})$$

Extreme Value Theorem

Let $C \subset \mathbb{R}^n$ and $f_0 : \mathbb{R}^n \to \mathbb{R}$ such that:

- f_0 is continuous (actually only lower semicontinuous) and
- C is a **closed** and
- C is bounded, **or** f_0 grows to $+\infty$ along every direction ($\lim_{\|\mathbf{x}\| \to \infty} f_0(\mathbf{x}) = \infty$).

Then, it exists $\mathbf{x}^* \in C$ such that $f_0(\mathbf{x}^*) = \inf_{\mathbf{x} \in C} f_0(\mathbf{x})$.

Let us review optimality conditions for unconstrained problems with differentiable objectives.

First-order necessary conditions for a minimum

Let $f : \mathbb{R}^n \to \mathbb{R}$ a continuously differentiable function in an open neighborhood of $\mathbf{x}^*$. Then, if $\mathbf{x}^*$ is a **local minimum** then

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

Second-order sufficient conditions for a minimum

Let $f : \mathbb{R}^n \to \mathbb{R}$ with Hessian $\nabla^2 f$ continuous in an open neighborhood of $\mathbf{x}^*$. Then if

$$\nabla f(\mathbf{x}^*) = 0 \text{ and } \nabla^2 f(\mathbf{x}^*) \text{ is positive definite,}$$

$\mathbf{x}^*$ is a **local minimum**.

Main idea: approximate f by its 2nd order Taylor polynomial around $\mathbf{x}^*$:

$$f(\mathbf{x}) \approx f(\mathbf{x}^*) + \frac{1}{2}(\mathbf{x} - \mathbf{x}^*)^T \nabla^2 f(\mathbf{x}^*)(\mathbf{x} - \mathbf{x}^*).$$

$$\nabla^2 f(\mathbf{x}^*) = \begin{pmatrix} f_{x_1,x_1}(\mathbf{x}^*) & f_{x_1,x_2}(\mathbf{x}^*) & \dots & f_{x_1,x_n}(\mathbf{x}^*) \\ f_{x_2,x_1}(\mathbf{x}^*) & f_{x_2,x_2}(\mathbf{x}^*) & \dots & f_{x_2,x_n}(\mathbf{x}^*) \\ \vdots & \vdots & \ddots & \vdots \\ f_{x_n,x_1}(\mathbf{x}^*) & f_{x_n,x_2}(\mathbf{x}^*) & \dots & f_{x_n,x_n}(\mathbf{x}^*) \end{pmatrix} \quad (n \times n \text{ sym. matrix})$$

Parenthesis: quadratic functions in $\mathbb{R}^n$

Let $\mathbf{A}$ $n \times n$ symmetric, $\mathbf{b} \in \mathbb{R}^n$ and $c \in \mathbb{R}$:

$$p(\mathbf{x}) = \frac{1}{2}\mathbf{x}^T\mathbf{A}\mathbf{x} + \mathbf{b}^T\mathbf{x} + c$$

We focus on the case in which $\mathbf{b} = 0$ and $c = 0$.

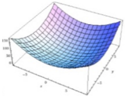
When $n = 1$**:** $p(x) = \frac{1}{2}ax^2$.

- $a > 0$: upwards parabola
- $a < 0$ downwards parabola.

For $n > 1$**:** we need to look at the eigenvalues of $\mathbf{A}$: $\lambda_1, \ldots, \lambda_n$. There are several possibilities.

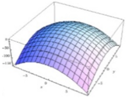
Parenthesis: quadratic functions in $\mathbb{R}^n$

$\lambda_i > 0,\ i = 1, \ldots, n$ **A** is **positive definite**



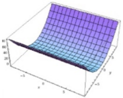
upwards paraboloid

$\lambda_i < 0,\ i = 1, \ldots, n$ **A** is **negative definite**



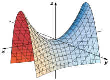
downwards paraboloid

$\lambda_i \geq 0,\ i = 1, \ldots, n$ **A** is **positive semidefinite**



upwards parabollic valey

$\lambda_i > 0, \lambda_j < 0$ **A** is **indefinite**



saddle

Second-order sufficient conditions for a minimum

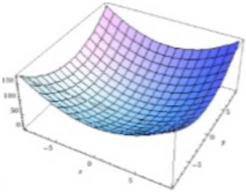
Let $f : \mathbb{R}^n \to \mathbb{R}$ with Hessian $\nabla^2 f$ continuous in an open neighborhood of $\mathbf{x}^*$. Then if

$$\nabla f(\mathbf{x}^*) = 0 \text{ and } \nabla^2 f(\mathbf{x}^*) \text{ is positive definite,}$$

$\mathbf{x}^*$ is a **local minimum**.

Main idea: If f around $\mathbf{x}^*$ looks like an upwards paraboloid, then $\mathbf{x}^*$ is a local minimum

$f(\mathbf{x}) \approx$



We want to find the global minimum of a function f over a set C.

In general, non-convex problems (when either f or C are non-convex) are hard:

- The problem can be intractale: no known solution polinomial time with the size of the problem.
- All we can hope to find is a local minimizer.

Convex problems are easy.

Convex constrained minimization

Consider the **constrained minimization problem**

$$\min_{\mathbf{x} \in C} f(\mathbf{x}).$$

Theorem

Assume that C is a **convex subset** of $\mathbb{R}^n$. Let $f : \mathbb{R}^n \to (-\infty, +\infty]$ be a **convex function**.

Then, **a local minimum of** f **over** C **is also a global minimum over** C.

Moreover, if f is **strictly convex**, then **the global minimum in** C **is unique**.

An optimization problem

$$\min_{\mathbf{x} \in C} f_0(\mathbf{x})$$

is **convex** if

- f_0 is a convex function
- C is a convex set

If can express our problem as a convex optim, we are done!

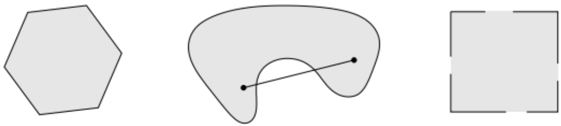
- Convex optimization problems are easy.
- Any local minimizer is a global minimizer.
- The problem can be solved in polynomial time, and there exist many efficient optimization toolboxes that work for most problems.
- Many convex problems can be considered a technology, analogous to solving a system of equations. Many efficient convex optimization solvers exist.
- Therefore it is important to distinguish convex problems from non-convex ones.

Convex sets

A set $C \in \mathbb{R}^N$ is **convex** if given two points $\mathbf{x}, \mathbf{y} \in C$, the segment joining $\mathbf{x}$ and $\mathbf{y}$ is contained in C. That is, given any two points $\mathbf{x}, \mathbf{y} \in C$, then

$$[\mathbf{x}, \mathbf{y}] := \{t\mathbf{x} + (1-t)\mathbf{y} : t \in [0,1]\} \subset C.$$

Examples:



From the course slides *Convex Optimization - Boyd & Vandenberghe*

Convex sets in $\mathbb{R}$

In $\mathbb{R}$ the convex sets are the intervals, for example for $a, b \in \mathbb{R}$, $a \leq b$:

- (a, b)
- $[a, b)$
- $[a, +\infty)$
- $(-\infty, a)$

Norm balls

A **norm ball** with center $\mathbf{x}_c \in \mathbb{R}^n$ and radius $r \in \mathbb{R}$ is the set

$$\{\mathbf{x} \in \mathbb{R}^n \mid \|\mathbf{x} - \mathbf{x}_c\| \leq r\}$$

Example: An **euclidean ball** with center $\mathbf{x}_c$ and radius r:

$$B_2(\mathbf{x}_c, r) = \{\mathbf{x} \mid \|\mathbf{x} - \mathbf{x}_c\|_2 \leq r\} = \left\{\mathbf{x} \mid \sqrt{\sum_{i=1}^{n} (x_i - x_{c,i})^2} \leq r\right\}$$

Example: An ℓ_1-**norm** with center $\mathbf{x}_c$ and radius r:

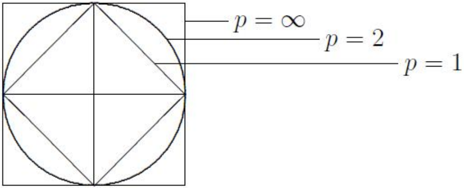
$$B_1(\mathbf{x}_c, r) = \{\mathbf{x} \mid \|\mathbf{x} - \mathbf{x}_c\|_1 \leq r\} = \left\{\mathbf{x} \mid \sum_{i=1}^{n} |x_i - x_{c,i}| \leq r\right\}$$

Example: An ℓ_∞-**norm** with center $\mathbf{x}_c$ and radius r:

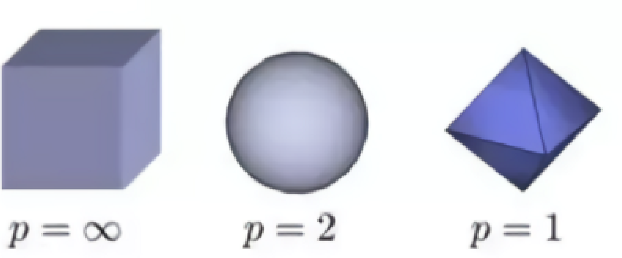
$$B_\infty(\mathbf{x}_c, r) = \{\mathbf{x} \mid \|\mathbf{x} - \mathbf{x}_c\|_\infty \leq r\} = \left\{\mathbf{x} \mid \max_{i=1,\ldots,n} |x_i - x_{c,i}| \leq r\right\}$$

Norm balls

In $\mathbb{R}^2$:



In $\mathbb{R}^3$:

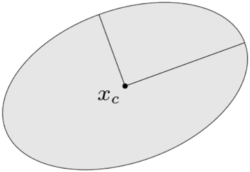


Norm balls - ellipsoides

An **ellipsoid** is a set of the form

$$\{\mathbf{x} | (\mathbf{x} - \mathbf{x}_c)^t \mathbf{A} (\mathbf{x} - \mathbf{x}_c) \leq 1\}$$

with A a symmetric positive definite matrix.



From the course slides *Convex Optimization - Boyd & Vandenberghe*

A set $S \in \mathbb{R}^N$ is **affine** if given two points $\mathbf{x}, \mathbf{y} \in S$, the line through $\mathbf{x}$ and $\mathbf{y}$ is contained in S. That is, given any two points $\mathbf{x}, \mathbf{y} \in S$, then

$$[\mathbf{x}, \mathbf{y}] := \{t\mathbf{x} + (1-t)\mathbf{y} : t \in \mathbb{R}\} \subset S.$$

Therefore, **every affine set is a convex set.**

Example: A solution set of linear equations $\{\mathbf{x} \mid \mathbf{A}\mathbf{x} = \mathbf{b}\}$.

Moreover, every affine set can be expressed as solution set of system of linear equations.

Lines in in $\mathbb{R}^2$:

$$\{\mathbf{x} = (x_1, x_2) \in \mathbb{R}^2 \,|\, a_1x_1 + a_2x_2 = b\},$$

where $\mathbf{a} = (a_1, a_2)$ is the normal to the line.

Planes in $\mathbb{R}^3$:

$$\{\mathbf{x} = (x_1, x_2, x_3) \in \mathbb{R}^3 \,|\, \mathbf{a}^T\mathbf{x} = a_1x_1 + a_2x_2 + a_3x_3 = b\},$$

where $\mathbf{a} = (a_1, a_2, a_3)$ is the normal to the line.

Lines in $\mathbb{R}^3$ (as intersection of two planes):

$$\left\{\mathbf{x} = (x_1, x_2, x_3) \in \mathbb{R}^3 \,|\, \mathbf{A}\mathbf{x} = \mathbf{b}\right\}$$

where

$$\mathbf{A} = \begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \end{pmatrix} \qquad \text{and} \qquad \mathbf{b} = \begin{pmatrix} b_1 \\ b_1 \end{pmatrix}.$$

The rows $\mathbf{a}_1, \mathbf{a}_2 \in \mathbb{R}^3$ of $\mathbf{A}$ are the normals to the planes.

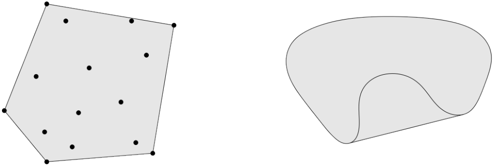
Convex combination and convex hull

A **convex combination** of $\mathbf{x}_1, \ldots, \mathbf{x}_k$ is any point $\mathbf{x}$ of the form

$$\mathbf{x} = t_1\mathbf{x}_1 + t_2\mathbf{x}_2 + \cdots + t_k\mathbf{x}_k$$

with $t_1 + t_2 + \cdots + t_k = 1,\ t_i \geq 0$.

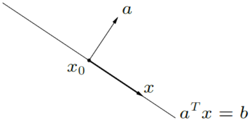
The **convex hull** of a set S is the set of all convex combinations of points in S.



From the course slides *Convex Optimization - Boyd & Vandenberghe*

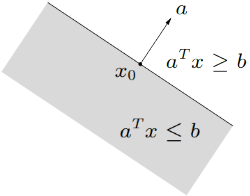
Hyperplanes and half-spaces

A **hyperplane** is a set of the form $\{\mathbf{x} \mid \mathbf{a}^t\mathbf{x} = b\}, (\mathbf{a} \neq 0)$



From the course slides *Convex Optimization - Boyd & Vandenberghe*

A **halfspace** is a set of the form $\{x \mid \mathbf{a}^t\mathbf{x} \leq b\}, (\mathbf{a} \neq 0)$



From the course slides *Convex Optimization - Boyd & Vandenberghe*

In both cases a is the normal vector.

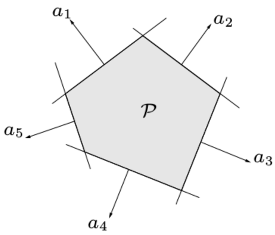
Polytope

A **polytope** is a set of the form

$$\{\mathbf{x} \mid \mathbf{A}\mathbf{x} \preceq \mathbf{b}\} \quad \mathbf{A} \in R^{m \times n}.$$

$\preceq$ is component-wise inequality: $\mathbf{a}_i\mathbf{x} \leq b_i$, where $\mathbf{a}_i$ is the i-th row of $\mathbf{A}$.

A polytope is an intersection of m half-spaces:



From the course slides *Convex Optimization - Boyd & Vandenberghe*

Operations that preserve convexity

The **intersection** of (any number of) convex sets is convex.

If $f : \mathbb{R}^n \to \mathbb{R}^n$ is an **affine function** (i.e. it can be written as $f(\mathbf{x}) = \mathbf{A}\mathbf{x} + \mathbf{b}$ with $\mathbf{A} \in M_{n,m}(\mathbb{R})$, $\mathbf{b} \in \mathbb{R}^m$), then:

- If $C \subseteq \mathbb{R}^n$ is a convex set, then $f(C) = \{f(\mathbf{x}) \in \mathbb{R}^m \mid \mathbf{x} \in C\}$ is a convex set.
- If $C \subseteq \mathbb{R}^m$ is a convex set, then $f^{-1}(C) = \{\mathbf{x} \in \mathbb{R}^n \mid f(\mathbf{x}) \in C\}$ is a convex set.

To verify if a set is convex, we almost never use the definition. Instead, we try to express in terms of simpler known convex sets using operations that preserve convexity

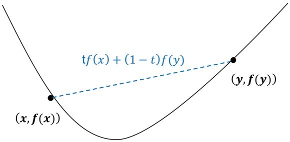
Convex functions

Convex Functions

A function $f : C \to \mathbb{R}$ is **convex** if

$$f(t\mathbf{x} + (1-t)\mathbf{y}) \leq tf(\mathbf{x}) + (1-t)f(\mathbf{y}),$$

for any two points $\mathbf{x}, \mathbf{y} \in C$ and any $t \in [0,1]$.



Conversely, f is concave if $-f$ is convex.

A function $f : C \to \mathbb{R}$ is **strictly convex** if

$$f(t\mathbf{x} + (1-t)\mathbf{y}) < tf(\mathbf{x}) + (1-t)f(\mathbf{y}),$$

for any two points $\mathbf{x}, \mathbf{y} \in C$ and any $t \in [0,1]$.

Examples of convex and concave functions on $\mathbb{R}$

Convex functions

- *Quadratic functions*: x^2 on $\mathbb{R}$.
- *Exponential*: e^{ax} is convex on $\mathbb{R}$, for any $a \in \mathbb{R}$.
- *Powers*: x^α for $x > 0$ and $\alpha \geq 1$ or $\alpha \leq 0$.
- *Powers of absolute value*: $|x|^p$ on $\mathbb{R}$, for $p \geq 1$.
- *Negative entropy*: $x \log x$ for $x > 0$.
- *Affine functions*: $ax + b$ on $\mathbb{R}$, for any $a, b \in \mathbb{R}$.

Concave functions

- *Affine functions*: $ax + b$ on $\mathbb{R}$, for any $a, b \in \mathbb{R}$. (Convex and concave!)
- *Powers*: x^α for $x > 0$ and $0 \leq \alpha \leq 1$.
- *Logarithm*: $\log x$ for $x > 0$.

Examples of convex functions on $\mathbb{R}^n$

- *Affine functions*: $f(\mathbf{x}) = \langle \mathbf{x}, \mathbf{b} \rangle + c$.
- *Norms*: Every norm on $\mathbb{R}^n$ is convex.

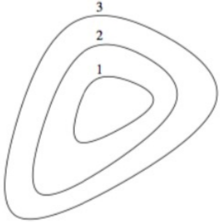
  In particular any ℓ_p norm, $\|\mathbf{x}\|_p = \left(\sum_{i=1}^{n} |x_i|^p \right)^{1/p}$ is convex.
- *Max function:* $f(\mathbf{x}) = \max\{x_1, \ldots, x_n\}$.
- *Quadratic-over-linear function*. $f(x_1, x_2) = x_1^2 / x_2$, with $x_2 > 0$.

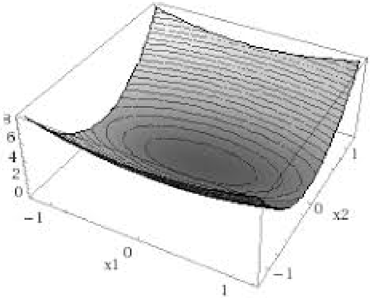
Epigraph and sub-levels

α-**sublevel set** of $f : \mathbb{R}^n \to \mathbb{R}$:

$$C_\alpha = \{\mathbf{x} \in \mathrm{dom}(f) \,|\, f(\mathbf{x}) \leq \alpha\}.$$

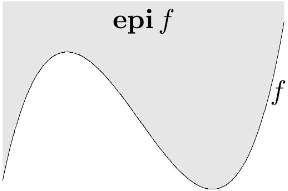
- sublevel sets of convex functions are convex.
- functions with all their sub-level sets convex, are not necessary convex. These broarder class of functions are called **quasi-convex** (and are also easy to optimize).





Epigraph and sub-levels

epigraph of $f : \mathbb{R}^n \to \mathbb{R}$: $\mathrm{epi}(f) = \{(\mathbf{x}, t) \in \mathbb{R}^{n+1} \,|\, \mathbf{x} \in \mathrm{dom}(f), f(\mathbf{x}) \leq t\}$.



f is a **convex function** if and only if its epigraph is a **convex set**

Sufficient and necessary condition for convexity

A function $f : \mathbb{R}^n \to \mathbb{R}$ is convex if and only if the function $g : \mathbb{R} \to \mathbb{R}$,

$$g(t) = f(\mathbf{x} + t\mathbf{v})$$

is convex for any $\mathbf{x}, \mathbf{v} \in \mathbb{R}^n$.

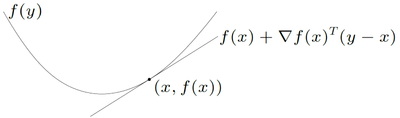
Using this property the problem is reduced to check the convexity of function of one variable.

First order condition for convexity

Let $f : \mathbb{R}^n \to \mathbb{R}^m$ be a **differentiable** function and let C be a convex subset of $\mathbb{R}^n$.

Then f is **convex** if and only if:

$$f(\mathbf{y}) \geq f(\mathbf{x}) + \langle \nabla f(\mathbf{x}), \mathbf{y} - \mathbf{x} \rangle, \quad \forall \mathbf{x}, \mathbf{y} \in C.$$



From the course slides *Convex Optimization - Boyd & Vandenberghe*

(*i.e.* iff the first order approximation of f is a global underestimator.)

Moreover, f is **strictly convex** if and only if:

$$f(\mathbf{y}) > f(\mathbf{x}) + \langle \nabla f(\mathbf{x}), \mathbf{y} - \mathbf{x} \rangle, \quad \forall \mathbf{x}, \mathbf{y} \in C.$$

Let $f : \mathbb{R}^n \to \mathbb{R}^m$ be a **twice differentiable** function and let C be a convex subset of $\mathbb{R}^n$.

Then f is **convex** if and only if the Hessian of f, $\nabla^2 f(\mathbf{x})$, is **positive semi-definite** for all points $\mathbf{x} \in C$. This is, all the eigenvalues of $\nabla^2 f(\mathbf{x})$ are non-negative.

Moreover, if $\nabla^2 f(\mathbf{x})$ is **positive definite** for all points $\mathbf{x} \in C$ (*i.e.* all its eigenvalues are positive) then f is **strictly convex**.

Is a quadratic function a convex function?

Let consider $f(\mathbf{x}) = \dfrac{1}{2}\langle \mathbf{x}, \mathbf{A}\mathbf{x}\rangle - \langle \mathbf{x}, \mathbf{b}\rangle + c$, with $\mathbf{A}$ a symmetric matrix.

Then, $\nabla f(\mathbf{x}) = \mathbf{A}\mathbf{x} - \mathbf{b}$ and $\nabla^2 f(\mathbf{x}) = \mathbf{A}$.

$\rightarrow$ $f(\mathbf{x})$ **is convex if the matrix A is positive semi-definite.**

Let consider $f(\mathbf{x}) = \dfrac{1}{2}\|\mathbf{A}x - \mathbf{b}\|_2^2$.

Then, $\nabla f(\mathbf{x}) = \mathbf{A}^t(\mathbf{A}\mathbf{x} - \mathbf{b})$ and $\nabla^2 f(\mathbf{x}) = \mathbf{A}^t\mathbf{A}$.

$\rightarrow$ $f(\mathbf{x})$ **is convex for any matrix A** because $\mathbf{A}^t\mathbf{A}$ is symmetric and positive semidefinite.

In practice, for a general function f we have to either

1. show that f is obtained from simple convex functions and operations that preserve convexity (next slide)
2. verify the definition (restricted to a line),
3. if the function is twice differentiable, show that $\nabla^2 f(\mathbf{x})$ is positive semidefinite,
4. verify the definition.

Operations that preserve convexity

Nonnegative multiple: af is convex if f is convex and $\alpha > 0$.

Sum: $f_1 + f_2$ is convex if f_1, f_2 are convex.
This is also true for infinite sums and for integrals.

Composition with affine function:
Let $f : \mathbb{R}^n \to \mathbb{R}$ be a convex function, $\mathbf{A} \in M_{n,m}(\mathbb{R})$, $\mathbf{b} \in \mathbb{R}^n$. Then, $f(\mathbf{Ax} + \mathbf{b})$ is a convex function.

Example: any norm of an affine function $f(x) = \|\mathbf{Ax} + \mathbf{b}\|$ is convex.

Pointwise maximum:
If $f_1, \ldots, f_m$ are convex, then $f(\mathbf{x}) = \max\{f_1(\mathbf{x}), \ldots, f_m(\mathbf{x})\}$ is convex.

Pointwise supremum:
If $f(\mathbf{x}, \mathbf{y})$ is convex in $\mathbf{x}$ for all $\mathbf{y}$, and C is an arbitrary set, then

$$g(\mathbf{x}) = \sup_{\mathbf{y} \in C} f(\mathbf{x}, \mathbf{y}) \text{ is convex.}$$

Pointwise infimum: If $f(\mathbf{x}, \mathbf{y})$ is convex in $(\mathbf{x}, \mathbf{y})$ and C is a convex set, then

$$g(\mathbf{x}) = \inf_{\mathbf{y} \in C} f(\mathbf{x}, \mathbf{y}) \text{ is convex.}$$

Convex Problems you have already seen...

... written as finite dimensional problems (i.e. as matrices and vectors).

Image denoising

Given $\mathbf{f}$ a noisy image, recover $\mathbf{u}$ as the solution of

$$\min_{\mathbf{u}} \|\nabla^h \mathbf{u}\|_2^2 + \lambda \|\mathbf{u} - \mathbf{f}\|_2^2$$

Image inpainting

Given $\mathbf{f}$ an image and a mask $\mathbf{m}$ defining the region that should be preserved:

$$\min_{\mathbf{u}} \|\nabla^h \mathbf{u}\|_2^2 \quad \text{s.t. } \mathbf{m} \odot \mathbf{u} = \mathbf{f}$$

Remember that approximating ∇ with finite differences it can be expressed as a matrix.

Total Variation Image denoising

Given $\mathbf{f}$ a noisy image, recover $\mathbf{u}$ as the solution of

$$\min_{\mathbf{u}} \|\nabla^h \mathbf{u}\|^1 + \lambda \|\mathbf{u} - \mathbf{f}\|_2^2$$

For working with this type of problems you should wait until next lecture.

What's next?

We will study **how to compute the minimum of a convex function with convex restrictions on its variables**.

The solutions will satisfy the so-called the **Karush-Kuhn-Tucker (KKT) optimality conditions**.

The KKT optimality conditions are the necessary and sufficient conditions of a minimum. They allow to write equations to compute the solution to the problems.

Extra material

Norms

A **norm** is a function $\|\cdot\|$ that satisfies

- $\|x\| \geq 0$ and $\|x\| = 0$ if and only if $x = 0$.
- $\|tx\| = |t|\|x\|$ for $t \in \mathbb{R}$
- $\|x+y\| \leq \|x\| + \|y\|$

Examples: If $x \in \mathbb{R}^n$ we already know the euclidean norm (or ℓ_2 norm):

$$\|x\|_2 = \sqrt{|x_1|^2 + \cdots + |x_n|^2} = \left(\sum_{i=1}^{n} |x_i|^2\right)^{1/2}$$

We should also be familiar with the ℓ_1 norm:

$$\|x\|_1 = |x_1| + \cdots + |x_n|$$

And in general we define the ℓ_p norm for $p \geq 1, p \in \mathbb{R}$:

$$\|x\|_p = \left(\sum_{i=1}^{p} |x_i|^p\right)^{1/p}$$