General optimization and convex analysis

Pierre Hellier

General formulation

General problem formulation for unconstrained optimization

Solve for $x^* = arg\min_{\mathbf{x} \in C} L(x)$

- L is the loss function
- $\mathbf{x} \in \mathbb{R}^n$ is a vector of n variables
- C is the set of admissible solutions
- Objective: find vector of minimal value $\mathbf{x}^*$, i.e. $\forall \mathbf{x} \in C, L(\mathbf{x}^*) < L(\mathbf{x})$

In the following, the loss, or function, will be denoted L or F

General problem formulation for constrained optimization

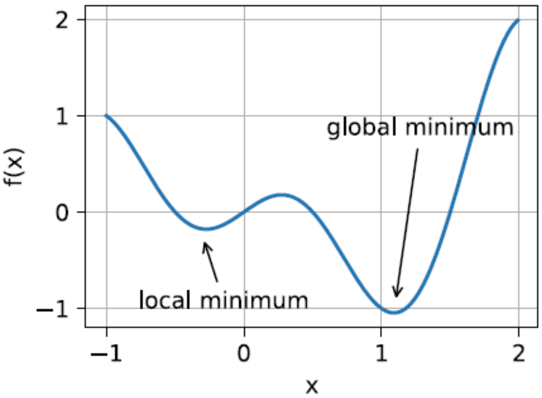
Solve for $x^* = arg \min_{\mathbf{x} \in \mathbb{R}^n} L(x)$
with $h_j(x) = 0, \forall j = 1, \ldots, p$
and $g_i(x) < 0, \forall i = 1, \ldots, q$

- h_j and g_i define equality and inequality constraints
- Equivalent to unconstrained optimization if we define $C = \{x \in \mathbb{R}^n | h_j(x) = 0, \forall j = 1, \ldots, p \text{ and } g_i(x) < 0, \forall i = 1, \ldots, q\}$
- if $p = q = 0$, equivalent to unconstrained optimization

Definitions

- **Feasible point** Any point $x \in C$ that satisfy the constraints
- **Optimal value** Minimal value function $L^* = L(x^*)$
- **Optimal solution** x^* is an optimal solution if $\forall x \in C, L(x^*) < L(x)$
- **Sub-optimal solution** x^* is an sub-optimal solution (local optimum) only if it is an optimal solution on a ball around x^*.

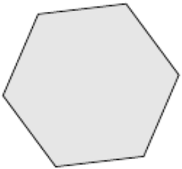
Minimas

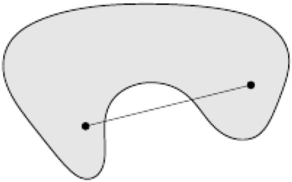


Convexity

Convexity

Convex sets are defined so that the line between two points is the set:
C is a convex set if $\forall x, y \in C, 0 < \alpha < 1, \alpha x + (1 - \alpha) y \in C$



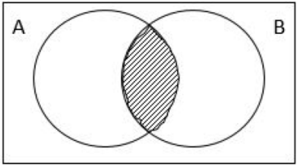


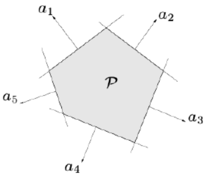
Examples of convex sets

- $\mathbb{R}^n$
- Positive orthant $\mathbb{R}^n_+$
- Hyperplan $\{x \in \mathbb{R}^n | a^T x = b\}$
- Half space $\{x \in \mathbb{R}^n | a^T x < b\}$
- Polyhedra $\{x \in \mathbb{R}^n | Ax < b\}$

Operations preserving convexity

Intersection if C_k are all convex, then $\cap_k C_k$ is convex





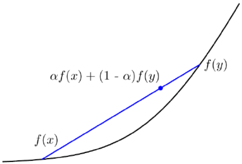
Operations preserving convexity

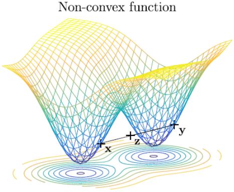
- Cartesian product $C_k \subseteq \mathbb{R}^{n_k}$ are all convex, then $C_1 \times C_2 \times \cdots \times C_M$ is convex
- Affine transform. if $C \subseteq \mathbb{R}^n$ is convex, and $\mathcal{A}$ is an affine transformation ($\mathcal{A}(x) = Ax + b$), then $\mathcal{A}(C) = \{\mathcal{A}(x), x \in C\}$ is convex

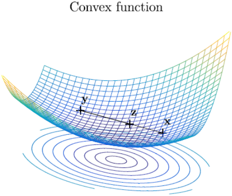
Convexity (1)

Convex functions: a function is convex if it lies below its chords

$$f \text{ convex } \equiv \forall x, y, \forall 0 < \alpha < 1, f(\alpha x + (1-\alpha) y) \leq \alpha f(x) + (1-\alpha) f(y)$$







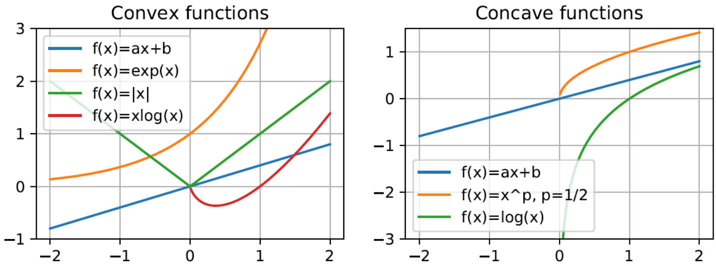
Convexity (2)

- Strictly convex: the inequality is strict
- if f is convex, $\{x|f(x) \leq 0\}$ is convex
- if f is convex, $-f$ is concave
- if f is twice differentiable, f convex $\equiv f^{''} > 0$

Convex functions examples

Some usual convex functions:

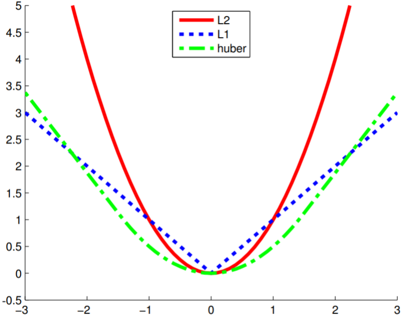
- affine functions
- exponential functions
- power functions (power > 1)
- neg-entropy $x \log(x), x > 0$

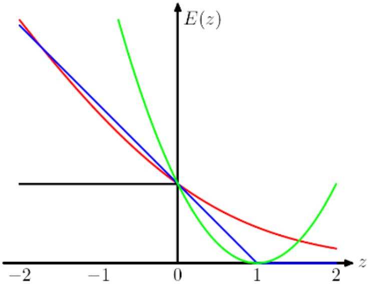


Operations preserving convexity

- Positive sum of two functions $f = \lambda_1 f_1 + \lambda_2 f_2$ with $\lambda > 0$
- Composition with affine function $f(Ax + b)$
- Composition of convex function g with convex increasing function h, $f(x) = h(g(x))$
- Maximum $f_1, \ldots, f_m$ are convex, $f(x) = \max_i\{f_i(x)\}$ is convex

Generally, loss functions are convex functions





Convexity and optimal solution

- Convex problem: optimization of convex funtion over a convex set
- A local minimum of a convex problem is a global minimum
- if strict convexity, only one unique global minimum

Smoothness and gradients

Differentiability

- function f is differentiability class C^k if $\frac{d^k f}{dx^k}$ is continuous
- C^0 are the continuous functions
- C^1 functions have continuous derivatives
- Smooth function is C^∞

Exercice: for each function, determine differentiability class, convexity and plot (with matplolib). Functions:
$2x + 1, x^2, e^x, |x|, max(x, 0), log(1 + e^x)$

Gradient

- F function $\mathbb{R}^n \to \mathbb{R}$
- Gradient: Vector of partial derivatives
- $\nabla_x F(x) = \{\frac{\partial F(x)}{\partial x_1}, \ldots, \frac{\partial F(x)}{\partial x_m}\}^T$
- Differentiability reads:
  $F(x+\epsilon) = F(x) + \langle \epsilon, \nabla_x F(x) \rangle + o(||\epsilon||)$ with $\lim_{\epsilon \to 0} o(||\epsilon||)/\epsilon = 0$
- Direction of gradient: steepest direction (where F increases the most)

Convexity and differentiability

relation

f convex $\Leftrightarrow \forall (x, x'), f(x) \geq f(x') + \langle \nabla_x F(x'), x - x' \rangle$

Gradient and optimality

Gradient is a necessary condition for optimality

if x^* is a local minimum of function f (on some ball around x^*), then $\nabla_x F(x^*) = 0$

Problem

Exercice: proove it!
Hint: Write $f(x^*)$ with a small deviation, perform first-order Taylor expansion, use opposite direction.

Solution

for small ϵ and u fixed:

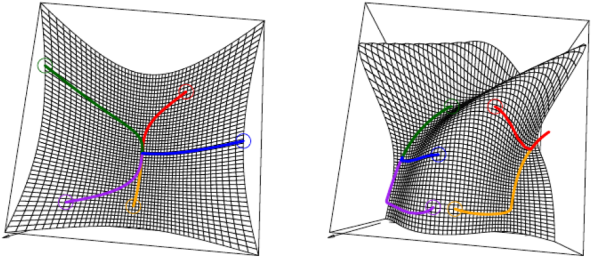
$$f(x^*) \leq f(x^*+\epsilon u) = f(x^*)+\epsilon\langle\nabla F(x^*), u\rangle+o(\epsilon) \rightarrow \langle\nabla F(x^*), u\rangle \geq 0$$

Apply the same for $-u$ leads to $\langle\nabla F(x^*), u\rangle = 0, \forall u$

Convexity and global optimum

If f is convex, local minima are global minima

$$x^* \in arg\min_x f(x) \Leftrightarrow \nabla f(x^*) = 0$$



Convex Nonconvex

Convexity and global optimum

Problem

Exercice:

- Prove that for a convex function, a local minimum is a global minimum
- Prove that for a convex function, a zero gradient is a sufficient condition of optimality

Solution (1)

For any x, there is a small t such that $tx^* + (1-t)x$ is close to x^*

$$f(x^*) \leq f(tx^* + (1-t)x) \leq tf(x^*) + (1-t)f(x) \Rightarrow f(x^*) \leq f(x)$$

and thus x^* is a global minimum

Solution (2)

We already know that if a point is a local minimum, its gradient is naught. Let us proof that this is a sufficient condition for convex functions.
Let us assume that $\nabla F(x^*) = 0$
By convexity:

$$f(x) \geq f(x^*) + \langle \nabla_x F(x^*), x - x^* \rangle = f(x^*)$$

and thus x^* is a global minimum