Existence of optimal solutions and optimality conditions

Mauro Passacantando

Department of Computer Science, University of Pisa
mauro.passacantando@unipi.it

Optimization Methods
Master of Science in Embedded Computing Systems – University of Pisa
http://pages.di.unipi.it/passacantando/om/OM.html

Optimization problem in standard form

$$\begin{cases} \min \ f(x) \\ g(x) \le 0 \\ h(x) = 0 \end{cases}$$

- $f : \mathbb{R}^n \to \mathbb{R}$ is the objective function
- $g_i : \mathbb{R}^n \to \mathbb{R}$, $i = 1, \dots, m$ are the inequality constraints functions
- $h_j : \mathbb{R}^n \to \mathbb{R}$, $j = 1, \dots, p$ are the equality constraints functions

Domain: $\mathcal{D} = \mathrm{dom}(f) \cap \bigcap_{i=1}^{m} \mathrm{dom}(g_i) \cap \bigcap_{j=1}^{p} \mathrm{dom}(h_j)$

Feasible region: $\Omega = \{x \in \mathcal{D} : \ g(x) \le 0, \ h(x) = 0\}$

implicit constraint: $x \in \mathcal{D}$
explicit constraints: $g(x) \le 0$, $h(x) = 0$

Global and local optima

Global optimal solution: a feasible point x^* s.t. $f(x^*) \leq f(x)$ for all $x \in \Omega$

Local optimal solution: a feasible point x^* s.t. $f(x^*) \leq f(x)$ for all $x \in \Omega \cap B(x^*, R)$ for some $R > 0$

Optimal value:

$$v^* = \inf\{f(x) : \ x \in \Omega\}$$

$v^* = -\infty$ if the problem is unbounded below
$v^* = +\infty$ if the problem is infeasible

Examples

- $f(x) = \log(x)$, $v^* = -\infty$, no optimal solution
- $f(x) = e^x$, $v^* = 0$, no optimal solution
- $f(x) = x \log(x)$, $v^* = -1/e$, $x^* = 1/e$ is global optimum
- $f(x) = x^3 - 3x$, $v^* = -\infty$, $x^* = 1$ is local optimum

Convex optimization problems

An optimization problem

$$\begin{cases} \min\ f(x) \\ g(x) \leq 0 \\ h(x) = 0 \end{cases}$$

is convex if f is convex, $g_1, \dots, g_m$ are convex and $h_1, \dots, h_p$ are affine.

Examples

a) $\begin{cases} \min\ x_1^2 + x_1 x_2 + 3x_2^2 + 4x_1 + 5x_2 \\ x_1^2 + x_2^2 - 4 \leq 0 \\ x_1 + x_2 - 2 = 0 \end{cases}$ is convex

b) $\begin{cases} \min\ x_1^2 + x_2^2 \\ x_1/(1 + x_2^2) \leq 0 \\ (x_1 + x_2)^2 = 0 \end{cases}$ is NOT convex, but it is equivalent to:

$$\begin{cases} \min\ x_1^2 + x_2^2 \\ x_1 \leq 0 \\ x_1 + x_2 = 0 \end{cases}$$ which is convex

Convex optimization problems

Why convex problems are important?

Theorem 1

In a convex optimization problem the feasible region is convex

Theorem 2

In a convex optimization problem any local optimal solution is a global one

Proof. Let x^* be a local optimum, i.e. there is $R > 0$ s.t.

$$f(x^*) \leq f(z) \qquad \forall\ z \in \Omega \cap B(x^*, R).$$

By contradiction, assume that x^* is not a global optimum, i.e., there is $y \in \Omega$ s.t. $f(y) < f(x^*)$. Take $\alpha \in (0, 1)$ s.t. $\alpha x^* + (1-\alpha)y \in B(x^*, R)$. Then we have

$$f(x^*) \leq f(\alpha x^* + (1-\alpha)y) \leq \alpha f(x^*) + (1-\alpha)f(y) < f(x^*),$$

which is impossible. $\square$

Existence of global optima

Theorem (Weierstrass)

If the objective function f is continuous and the feasible region Ω is closed and bounded, then there exists a global optimum.

Proof. Let $v^* = \inf_{x \in \Omega} f(x)$. Define a minimizing sequence $\{x^k\} \subseteq \Omega$ s.t. $f(x^k) \to v^*$. Since $\{x^k\}$ is bounded, the Bolzano-Weierstrass theorem guarantees that there exists a subsequence $\{x^{k_p}\}$ converging to some point x^*. Since Ω is closed, we get $x^* \in \Omega$. Finally, $f(x^{k_p}) \to f(x^*)$ since f is continuous. Therefore, $f(x^*) = v^*$, i.e., x^* is a global optimum. $\square$

Corollary

If all the functions f, g_i, h_j are continuous, the domain $\mathcal{D}$ is closed and the feasible region Ω is bounded, then there exists a global optimum.

Example

$$\begin{cases} \min \ x_1 + x_2 \\ x_1^2 + x_2^2 - 4 \leq 0 \end{cases}$$

admits a global optimum. Where?

Existence of global optima

Corollary

If the objective function f is continuous, the feasible region Ω is closed and there exists $\alpha \in \mathbb{R}$ such that the α-sublevel set

$$\{x \in \Omega : \ f(x) \leq \alpha\}$$

is nonempty and bounded, then there exists a global optimum.

Example

$$\begin{cases} \min \ e^{x_1+x_2} \\ x_1 - x_2 \leq 0 \\ -2x_1 + x_2 \leq 0 \end{cases}$$

Ω is closed and unbounded. The sublevel set $\{x \in \Omega : \ f(x) \leq 2\}$ is nonempty and bounded, thus there exists a global optimum.

Existence of global optima

Corollary

If the objective function f is continuous and coercive, i.e.,

$$\lim_{\|x\| \to \infty} f(x) = +\infty,$$

and the feasible region Ω is closed, then there exists a global optimum.

Example

$$\begin{cases} \min \ x^4 + 3x^3 - 5x^2 + x - 2 \\ x \in \mathbb{R} \end{cases}$$

Since f is coercive, there exists a global optimum.

Existence and uniqueness of global optima

Corollary

- If f is strongly convex and Ω is closed, then there exists a global optimum.
- If f is strongly convex and Ω is closed and convex, then there exists a unique global optimum.

Example. Any quadratic programming problem

$$\begin{cases} \min \ \frac{1}{2}x^{\mathsf{T}}Qx + c^{\mathsf{T}}x \\ Ax \le b \end{cases}$$

where Q is a positive definite matrix has a unique global minimum.

What if Q is positive semidefinite or indefinite?

Existence of global optima for quadratic programming problems

Consider

$$\begin{cases} \min \ \frac{1}{2}x^{\mathsf{T}}Qx + c^{\mathsf{T}}x \\ Ax \leq b \end{cases} \tag{P}$$

The recession cone of Ω is $\text{rec}(\Omega) = \{d : \ Ad \leq 0\}$.

Theorem (Eaves)

(P) has a global optimal solution if and only if the following conditions hold:

(a) $d^{\mathsf{T}} Q\, d \geq 0$ for any $d \in \text{rec}(\Omega)$;

(b) $d^{\mathsf{T}}(Qx + c) \geq 0$ for any $x \in \Omega$ and any $d \in \text{rec}(\Omega)$ s.t. $d^{\mathsf{T}} Q\, d = 0$.

Existence of global optima for quadratic programming problems

Special cases:

- If $Q = 0$ (i.e., linear programming) then
  (P) has an optimal solution iff $d^{\mathsf{T}} c \geq 0 \ \forall \ d \in \mathrm{rec}(\Omega)$
- If Q is positive definite then (a) and (b) are satisfied.
- If Ω is bounded then (a) and (b) are satisfied.

Exercise. Prove that the quadratic programming problem

$$\begin{cases} \min \ \dfrac{1}{2}x_1^2 - \dfrac{1}{2}x_2^2 + x_1 - 2\,x_2 \\ -x_1 + x_2 \leq -1 \\ -x_2 \leq 0 \end{cases}$$

has a global optimal solution.

Unconstrained problems

Consider $\min_{x \in \mathbb{R}^n} f(x)$

Theorem (Necessary optimality condition)

If x^* is a local optimal solution, then

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

Proof. By contradiction, assume that $\nabla f(x^*) \neq 0$. Choose direction $d = -\nabla f(x^*)$, define $\varphi(t) = f(x^* + td)$,

$$\varphi'(0) = d^{\mathsf{T}} \nabla f(x^*) = -\|\nabla f(x^*)\|^2 < 0,$$

thus $f(x^* + td) < f(x^*)$ for all t small enough, which is impossible because x^* is a local optimum. □

Optimality condition for unconstrained convex problems

If f is convex, then x^* is a global minimum if and only if $\nabla f(x^*) = 0$.

Constrained problems

Example.

$$\begin{cases} \min \; x_1 + x_2 \\ x_1^2 + x_2^2 - 4 \leq 0 \end{cases}$$

$\Omega = B(0,2)$, global optimum is $x^* = (-\sqrt{2}, -\sqrt{2})$, $\nabla f(x^*) = (1,1)$.

Definition – Tangent cone

$$T_\Omega(x) = \left\{ d \in \mathbb{R}^n : \exists \; \{z_k\} \subset \Omega, \; \exists \; \{t_k\} > 0, \; z_k \to x, \; t_k \to 0, \; \lim_{k\to\infty} \frac{z_k - x}{t_k} = d \right\}$$

Example (continued). What is $T_\Omega(x^*)$?

First order necessary optimality condition

Theorem

If x^* is a local optimal solution, then

$$d^{\mathsf{T}} \nabla f(x^*) \geq 0, \qquad \forall\ d \in T_\Omega(x^*).$$

Proof. By contradiction, assume that there exists $d \in T_\Omega(x^*)$ s.t. $d^{\mathsf{T}} \nabla f(x^*) < 0$. Take the sequences $\{z_k\}$ and $\{t_k\}$ s.t. $\lim\limits_{k \to \infty} (z_k - x^*)/t_k = d$. Then $z_k = x^* + t_k\, d + o(t_k)$, where $o(t_k)/t_k \to 0$. The first order approximation of f gives

$$f(z_k) = f(x^*) + t_k\, d^{\mathsf{T}} \nabla f(x^*) + o(t_k),$$

thus there is $\bar{k} \in \mathbb{N}$ s.t.

$$\frac{f(z_k) - f(x^*)}{t_k} = d^{\mathsf{T}} \nabla f(x^*) + \frac{o(t_k)}{t_k} < 0 \qquad \forall\ k > \bar{k},$$

i.e. $f(z_k) < f(x^*)$ for all $k > \bar{k}$, which is impossible because x^* is a local optimum. $\square$

First order optimality condition for convex problems

Theorem

If Ω is convex, then $\Omega \subseteq T_\Omega(x) + x$ for any $x \in \Omega$.

Optimality condition for constrained convex problems

If the optimization problem is convex, then x^* is a global optimal solution if and only if

$$(y - x^*)^\mathsf{T} \nabla f(x^*) \geq 0, \qquad \forall\, y \in \Omega.$$

Exercise. Prove the latter result.

Properties of the tangent cone

$T_\Omega(x)$ is related to geometric properties of Ω.

Which is the relation between $T_\Omega(x)$ and constraints g, h defining Ω?

Example (continued). $g(x) = x_1^2 + x_2^2 - 4$, $\nabla g(x^*) = (-2\sqrt{2}, -2\sqrt{2})$,

$$T_\Omega(x^*) = \{d \in \mathbb{R}^2 : \ d^\mathsf{T} \nabla g(x^*) \leq 0\}$$

Definition – First-order feasible direction cone

Given $x \in \Omega$, $\mathcal{A}(x) = \{i : \ g_i(x) = 0\}$ denotes the set of inequality constraints which are active at x. The set

$$D(x) = \left\{ d \in \mathbb{R}^n : \quad \begin{array}{ll} d^\mathsf{T} \nabla g_i(x) \leq 0 & \forall\ i \in \mathcal{A}(x), \\ d^\mathsf{T} \nabla h_j(x) = 0 & \forall\ j = 1, \dots, p \end{array} \right\}$$

is called the *first-order feasible direction cone* at point x.

Properties of the tangent cone

Theorem

$T_\Omega(x) \subseteq D(x)$ for all $x \in \Omega$.

Definition – Abadie Constraints Qualification (ACQ)

Abadie Constraints Qualification holds at x when $T_\Omega(x) = D(x)$.

Remark

In general, ACQ does not hold at any $x \in \Omega$.

Example

$$\begin{cases} \min \ x_1 + x_2 \\ (x_1 - 1)^2 + (x_2 - 1)^2 - 1 \leq 0 \\ x_2 \leq 0 \end{cases}$$

$\Omega = \{(1,0)\}, \ T_\Omega(1,0) = \{(0,0)\}$.

$\nabla g_1(1,0) = (0,-2), \ \nabla g_2(1,0) = (0,1), \ D(1,0) = \{d \in \mathbb{R}^2 : \ d_2 = 0\}$.

Properties of the tangent cone

Theorem - Sufficient conditions for ACQ

a) (*Affine constraints*)
If g_i and h_j are affine for all $i = 1, \dots, m$ and $j = 1, \dots, p$, then ACQ holds at any $x \in \Omega$.

b) (*Slater condition*)
If g_i are convex for all $i = 1, \dots, m$, h_j are affine for all $j = 1, \dots, p$ and there exists $\bar{x} \in \text{int}(\mathcal{D})$ s.t. $g(\bar{x}) < 0$ and $h(\bar{x}) = 0$, then ACQ holds at any $x \in \Omega$.

c) (*Linear independence of the gradients of active constraints*)
If $\bar{x} \in \Omega$ and the vectors

$$\begin{cases} \nabla g_i(\bar{x}) & \text{for } i \in \mathcal{A}(\bar{x}), \\ \nabla h_j(\bar{x}) & \text{for } j = 1, \dots, p \end{cases}$$

are linear independent, then ACQ holds at any $\bar{x}$.

Karush-Kuhn-Tucker Theorem

Why ACQ is important?

Karush-Kuhn-Tucker Theorem

If x^* is a local optimum and ACQ holds at x^*, then there exist $\lambda^* \in \mathbb{R}^m$ and $\mu^* \in \mathbb{R}^p$ s.t. (x^*, λ^*, μ^*) satisfies the KKT system:

$$\begin{cases} \nabla f(x^*) + \sum_{i=1}^{m} \lambda_i^* \nabla g_i(x^*) + \sum_{j=1}^{p} \mu_j^* \nabla h_j(x^*) = 0 \\ \lambda_i^* g_i(x^*) = 0 \qquad \forall\ i = 1, \ldots, m \\ \lambda^* \geq 0 \\ g(x^*) \leq 0 \\ h(x^*) = 0 \end{cases}$$

Exercise. Use KKT system to solve

$$\begin{cases} \min\ x_1 - x_2 \\ x_1^2 + x_2^2 - 2 \leq 0 \end{cases}$$

Karush-Kuhn-Tucker Theorem

Remark

ACQ assumption is crucial in the KKT Theorem.

Example.

$$\begin{cases} \min x_1 + x_2 \\ (x_1 - 1)^2 + (x_2 - 1)^2 - 1 \leq 0 \\ x_2 \leq 0 \end{cases}$$

$x^* = (1, 0)$ is the global optimum.

$T_\Omega(x^*) = \{0\}$, $D(x^*) = \{d \in \mathbb{R}^2 : \ d_2 = 0\}$, hence ACQ does not hold at x^*.

$\nabla g_1(x^*) = (0, -2)$, $\nabla g_2(x^*) = (0, 1)$, $\nabla f(x^*) = (1, 1)$, hence there is no λ^* s.t. (x^*, λ^*) solves KKT system.

Karush-Kuhn-Tucker Theorem

KKT Theorem gives necessary optimality conditions, but not sufficient ones.

Example.

$$\begin{cases} \min x_1 + x_2 \\ -x_1^2 - x_2^2 + 2 \leq 0 \end{cases}$$

$x^* = (1,1)$, $\lambda^* = \dfrac{1}{2}$ solves KKT system, but x^* is not a local optimum.

KKT Theorem for convex problems

If the optimization problem is convex and (x^*, λ^*, μ^*) solves KKT system, then x^* is a global optimum.

Exercise. Prove the latter result.

Karush-Kuhn-Tucker Theorem

Exercise 1. Compute the distance between a point $z \in \mathbb{R}^n$ and the hyperplane $\{x \in \mathbb{R}^n : \ a^\mathsf{T} x = b\}$

Exercise 2. Compute the distance between two parallel hyperplanes

$$\{x \in \mathbb{R}^n : \ a^\mathsf{T} x = b_1\}, \qquad \{x \in \mathbb{R}^n : \ a^\mathsf{T} x = b_2\}, \qquad b_1 \neq b_2.$$

Exercise 3. Compute the projection of a point $z \in \mathbb{R}^n$ on the ball with center x^0 and radius r.

Exercise 4. Compute the projection of a point $z \in \mathbb{R}^2$ on the box

$$\{x \in \mathbb{R}^2 : \ a_1 \leq x_1 \leq b_1, \quad a_2 \leq x_2 \leq b_2\}.$$

Critical cone

Consider now a nonconvex optimization problem.

(x^*, λ^*, μ^*) solves KKT system. Is x^* a local optimum?

Definition – Critical cone

(x^*, λ^*, μ^*) solves KKT system. The critical cone is

$$C(x^*, \lambda^*, \mu^*) = \left\{ d \in \mathbb{R}^n : \quad \begin{array}{ll} d^{\mathsf{T}} \nabla g_i(x^*) = 0 & \forall\ i \in \mathcal{A}(x^*) \text{ con } \lambda_i^* > 0 \\ d^{\mathsf{T}} \nabla g_i(x^*) \leq 0 & \forall\ i \in \mathcal{A}(x^*) \text{ con } \lambda_i^* = 0 \\ d^{\mathsf{T}} \nabla h_j(x^*) = 0 & \forall\ j = 1, \dots, p \end{array} \right\}$$

Equivalent definition

$$C(x^*, \lambda^*, \mu^*) = \{ d \in D(x^*) : \ d^{\mathsf{T}} \nabla f(x^*) = 0 \}$$

Second order necessary optimality condition

Lagrangian function is defined as

$$L(x, \lambda, \mu) := f(x) + \sum_{i=1}^{m} \lambda_i \, g_i(x) + \sum_{j=1}^{p} \mu_j \, h_j(x)$$

Necessary condition

Assume that (x^*, λ^*, μ^*) solves KKT system and the gradients of active constraints at x^* are linear independent.
If x^* is a local optimum, then

$$d^{\mathsf{T}} \, \nabla^2_{xx} L(x^*, \lambda^*, \mu^*) \, d \geq 0 \qquad \forall \; d \in C(x^*, \lambda^*, \mu^*).$$

Special case: unconstrained problems

If x^* is a local optimum, then $\nabla^2 f(x^*)$ is positive semidefinite.

Second order necessary optimality condition

The previous theorem does not give a sufficient optimality condition.

Example.

$$\begin{cases} \min \ x_1^3 + x_2 \\ -x_2 \le 0 \end{cases}$$

$x^* = (0,0)$, $\lambda^* = 1$ is the unique solution of KKT system.
The linear constraint is active at x^* and $\nabla g(x^*) = (0,-1) \neq 0$.
Matrix $\nabla^2_{xx} L(x^*, \lambda^*) = 0$, but x^* is not a local optimum because $f(t,0) < f(0,0)$ for all $t < 0$.

Second order sufficient optimality condition

Sufficient condition

Assume that (x^*, λ^*, μ^*) solves KKT system and

$$d^{\mathsf{T}} \nabla^2_{xx} L(x^*, \lambda^*, \mu^*)\, d > 0 \qquad \forall\ d \in C(x^*, \lambda^*, \mu^*), d \neq 0,$$

then x^* is a local optimum.

Special case: unconstrained problems.

If $\nabla f(x^*) = 0$ and $\nabla^2 f(x^*)$ is positive definite then x^* is a local optimum.

Second order optimality conditions

Exercise. Find local and global optima of the following problems:

a) $\begin{cases} \min \ -2x_2^3 + x_1\, x_2^2 + x_1^2 - 2\, x_1\, x_2 + 3\, x_2^2 \\ x \in \mathbb{R}^2 \end{cases}$

b) $\begin{cases} \min \ -x_1^2 - 2\, x_2^2 \\ -x_1 + 1 \leq 0 \\ -x_2 + 1 \leq 0 \\ x_1 + x_2 - 6 \leq 0 \end{cases}$

c) $\begin{cases} \min \ -x_1 + x_2^2 \\ -x_1^2 - x_2^2 + 4 \leq 0 \end{cases}$

d) $\begin{cases} \min \ x_1^3 + x_2^3 \\ -x_1 - 1 \leq 0 \\ -x_2 - 1 \leq 0 \end{cases}$