2 - Existence of optimal solutions and optimality conditions for unconstrained problems

G. Mastroeni

Department of Computer Science, University of Pisa
giandomenico.mastroeni@unipi.it

Optimization Methods and Game Theory
Master of Science in Artificial Intelligence and Data Engineering
University of Pisa – A.Y. 2023/24

Contents of the lessons

- Existence of optimal solutions for optimization problems
- Existence of optimal solutions in the presence of convexity assumptions
- First and second order optimality conditions for unconstrained optimization problems
- Optimal solutions of unconstrained quadratic programming problems

Optimization problem in standard form

$$f_* = \min\{f(x) : \; x \in X\} \tag{P}$$

- $f : \mathbb{R}^n \to \mathbb{R}$ is the objective function
- $X \subseteq \mathbb{R}^n$ is the constraints set or feasible region
- If $X \equiv \mathbb{R}^n$ then (P) is said to be unconstrained

From now on, we will only consider minimization problems since

$$\max\{f(x) : \; x \in X\} = -\min\{-f(x) : \; x \in X\}.$$

In general X is defined by the expression:

$$X := \{x \in \mathbb{R}^n : g_i(x) \leq 0, \ i \in \mathcal{I}, \ h_j(x) = 0, \ j \in \mathcal{J}\},$$

where $g_i : \mathbb{R}^n \longrightarrow \mathbb{R}$, $i \in \mathcal{I} := \{1, .., m\}$, $h_j : \mathbb{R}^n \longrightarrow \mathbb{R}$, $j \in \mathcal{J} := \{1, .., p\}$,
Using the notation:

$$g(x) := (g_1(x), .., g_m(x))^T, \quad h(x) := (h_1(x), .., h_p(x))^T,$$

then $X = \{x \in \mathbb{R}^n : g(x) \leq 0, \ h(x) = 0\}$, $g : \mathbb{R}^n \longrightarrow \mathbb{R}^m$, $h : \mathbb{R}^n \longrightarrow \mathbb{R}^p$.

g and h are called the "constraint functions".

Remark

In this case (P) is also referred to as a *mathematical programming problem*.

Optimal value

The optimal value of (P) is defined by: $v(P) = \inf\{f(x) : \; x \in X\}$
$v(P) \in \mathbb{R}$ if the problem is bounded from below
$v(P) = -\infty$ if the problem is unbounded from below
$v(P) = +\infty$ if the problem is infeasible, i.e., $X = \emptyset$

Global optimal solution

A global optimal solution of (P) is a point $x^* \in X$ s.t. $f(x^*) \leq f(x)$ for all $x \in X$.
$X_* = \arg\min\{f(x) : \; x \in X\}$ denotes the set of global minima of f on X.

Local optimal solution

A local optimal solution of (P) is a point $x^* \in X$ s.t. $f(x^*) \leq f(x)$ for all $x \in X \cap B(x^*, r)$ for some $r > 0$.

Examples

- $f(x) = x$, $X = \mathbb{R}$, $v(P) = -\infty$, no optimal solution
- $f(x) = e^x$, $X = \mathbb{R}$, $v(P) = 0$, no optimal solution
- $f(x) = x^3 - 3x$, $X = \mathbb{R}$, $v(P) = -\infty$, $x^* = 1$ is a local optimum, no global optimum
- $f(x) = x \log(x)$, $X = \mathbb{R}_{++}$, $v(P) = -1/e$, $x^* = 1/e$ is a global optimum
- $f(x) = 3x^4 - 8x^3 - 6x^2 + 24x + 19$, $X = \mathbb{R}$, $v(P) = 0$, $x^* = -1$ is a global optimum and $\tilde{x} = 2$ is a local optimum

Theorem (Weierstrass)

If the objective function f is continuous and the feasible region X is closed and bounded, then (at least) a global optimum exists.

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

Example

$$\min \; x_1 + x_2 : \quad x \in X := \{(x_1, x_2) : x_1^2 + x_2^2 - 4 \leq 0\}$$

admits a global optimum.

Corollary 2

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

$$S_k(f) = \{x \in X : \ f(x) \leq k\} \tag{1}$$

is nonempty and bounded, then (at least) a global optimum exists.

Proof. Minimizing f on X is equivalent to minimize f on $S_k(f)$ which is bounded and closed since f is continuous and X is closed. □

Example

The function $f(x_1, x_2) = x_1^2 + x_2^2 - 4x_1 - 2x_2$ fulfils the condition (1).

In fact, suitably choosing k, the set

$$x_1^2 + x_2^2 - 4x_1 - 2x_2 \leq k$$

is a circle with center $C = (2, 1)$ and ray $r = \sqrt{2^2 + 1^2 + k}$

Example

$$\begin{cases} \min \ e^{x_1+x_2} \\ x \in X := \{x_1 - x_2 \leq 0, -2x_1 + x_2 \leq 0\} \end{cases}$$

f is continuous, X is closed and unbounded. But the sublevel set $S_2(f) = \{x \in X : \ f(x) \leq 2\}$ is nonempty and bounded, thus a global optimum exists.

Note that $S_2(f)$ is the solution set of the system:

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

Corollary 3

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

$$\lim_{\substack{\|x\|\to\infty \\ x\in X}} f(x) = +\infty, \tag{2}$$

and the feasible region $X \neq \emptyset$ is closed, then (at least) a global optimum exists.

Proof. Let $\bar{x} \in X$ and $k := f(\bar{x})$. By (2) the sublevel set $S_k(f)$ is nonempty and bounded, then apply Corollary 2 to show that a global optimum exists. □

Example

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

Since f is coercive, then there exists a global optimum.

Example

The function $f(x_1, x_2) = x_1^2 + x_2^2 - 4x_1 - 2x_2$ fulfils the condition (2), with $X = \mathbb{R}^2$.

Indeed, $f(x_1, x_2) = x_1^2 + x_2^2 - 4x_1 - 2x_2 = (x_1 - 2)^2 - 4 + (x_2 - 1)^2 - 1$.

If $\|x\| = \|(x_1, x_2)\| \to +\infty$, then at least one of the following conditions holds:

$$|x_1| \to +\infty \quad \textit{or} \quad |x_2| \to +\infty$$

In the first case we have:

$$\lim_{\|x\| \to +\infty} f(x_1, x_2) \geq \lim_{|x_1| \to +\infty} (x_1 - 2)^2 - 5 = +\infty$$

In the second case:

$$\lim_{\|x\| \to +\infty} f(x_1, x_2) \geq \lim_{|x_2| \to +\infty} (x_2 - 1)^2 - 5 = +\infty$$

Therefore, for any sequence $\{x^n\} = \{(x_1^n, x_2^n)\}$ such that $\|x^n\| \to +\infty$, we get

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

which guarantees that (2) is fulfilled.

Theorem 1

Assume that f is convex on the convex set X. Then any local optimum of (P) is a global optimum.

Proof. Let x^* be a local optimum of (P), i.e., there is $r > 0$ s.t.

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

By contradiction, assume that x^* is not a global optimum, then there exists $y \in X$ 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

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

that is impossible. □

Proposition 1

Assume that f is strictly convex on the convex set X and that (P) admits a global optimum x^*. Then x^* is the unique optimal solution of (P).

Proof. By contradiction, assume that there exists $\hat{x} \in X$, with $\hat{x} \neq x^*$, such that $f(\hat{x}) = f(x^*)$. Since f is strictly convex, we have

$$f(\alpha x^* + (1-\alpha)\hat{x}) < \alpha f(x^*) + (1-\alpha)f(\hat{x}) = f(x^*) \quad \forall \alpha \in (0,1)$$

□

which contradicts that x^* is a global optimum of (P). □

Theorem 2

If $f : \mathbb{R}^n \to \mathbb{R}$ is strongly convex on $\mathbb{R}^n$ and X is closed, then there exists a global optimum.

Proof. It is known that any convex function on $\mathbb{R}^n$ is continuous, moreover, recall that any strongly convex function is the sum of a convex function plus the function $\tau\|x\|_2^2$, for some $\tau > 0$, i.e.,

$$f(x) = \psi(x) + \tau\|x\|_2^2,$$

with ψ convex.
Since ψ is convex then it is bounded from below by linear function:

$$f(x) = \psi(x) + \tau\|x\|_2^2 \geq a^T x + \tau\|x\|_2^2 \geq -\|a\|_2\|x\|_2 + \tau\|x\|_2^2$$

By the previous inequalities it follows that f is coercive, so that the thesis follows from Corollary 3. □

Corollary 1

If f is strongly convex (on $\mathbb{R}^n$) and X is closed and convex, then there exists a unique global optimum.

Proof. By the previous theorem we know that a global minimum point exists, then the proof follows from Proposition 1. □

Example

Any quadratic programming problem

$$\min \ \frac{1}{2}x^\mathsf{T} Q x + c^\mathsf{T} x, \quad x \in X,$$

where Q is a positive definite matrix and X is closed and convex, has a unique global optimum.

We now consider the particular case where X is an open set. In particular, this assumption is fulfilled in when $X := \mathbb{R}^n$, i.e., (P) is an unconstrained problem defined by

$$\min\{f(x) : \ x \in \mathbb{R}^n\}.$$

Theorem 3 (Necessary optimality condition)

Assume that X is an open set and let f be differentiable at $x^* \in X$. If x^* is a local optimum of (P), 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. $\square$

Assume that $f : \mathbb{R}^n \longrightarrow \mathbb{R}$ has continuous first and second order partial derivatives for every $x \in X$.

Theorem 4 (Second order necessary optimality condition)

Let X be an open set and let $x^* \in X$ be a local optimum for (P). Then the following conditions hold:

- $\nabla f(x^*) = 0$;
- The Hessian matrix $\nabla^2 f(x^*)$ is positive semidefinite.

Theorem 5 (Second order sufficient optimality condition)

Let X be an open set, $x^* \in X$ and assume that the following conditions hold:

- $\nabla f(x^*) = 0$;
- The Hessian matrix $\nabla^2 f(x^*)$ is positive definite.

Then x^* is a local optimum for (P).

Theorem 6 (Optimality condition for convex problems)

Let f be a differentiable convex function on the open convex set X, then $x^* \in X$ is a global optimum for (P) if and only if $\nabla f(x^*) = 0$.

Proof. The necessity follows from Theorem 3.
Assume that $\nabla f(x^*) = 0$. Recall that, under the given differentiability assumptions f is convex on X if and only if

$$f(x) - f(y) \geq (x - y)^T \nabla f(y) \qquad \forall\, x, y \in X.$$

Setting $y = x^*$ we obtain

$$f(x^*) \leq f(x), \qquad \forall x \in X.$$

$\square$

Similarly we can prove the following uniqueness result.

Theorem 7

Let f be a differentiable strictly convex function on the open convex set X, then $x^* \in X$ is a unique global optimum for (P) if and only if $\nabla f(x^*) = 0$.

Existence of global optima for unconstrained quadratic programming problems

Consider the quadratic problem

$$\begin{cases} \min \; f(x) := \dfrac{1}{2}x^{\mathsf{T}}Qx + c^{\mathsf{T}}x \\ x \in \mathbb{R}^n \end{cases} \tag{P}$$

where Q is a $n \times n$ symmetric matrix.

Corollary 2

There exists a global optimum x^* for (P) if and only if the following conditions hold:

(i) $Qx^* + c = 0$,

(ii) Q is positive semidefinite.

Remark

Notice that, from (ii) a quadratic unconstrained problem admits an optimal solution only if f is convex, so that any local solution is also global.

Remark

We already observed that if Q is positive definite then (P) admits a unique global optimum. Indeed, in such a case Q is nonsingular and the system in (i) admits the unique solution $x^* = -Q^{-1}c$.

Let us consider more in details the case where Q is positive semidefinite but not positive definite.

In order to guarantee the existence of a global optimal solution we have to analyze the existence of a solution of the system $Qx + c = 0$.

By the Rouche'-Capelli Theorem the system $Qx = -c$ admits a solution if and only if

$$rank([Q, -c]) = rank(Q) \tag{3}$$

Proposition

If Q is positive semidefinite and (3) is fulfilled then (P) admits global optima given by the set of solutions of the system $Qx = -c$.

Example

Check if the function

$$f(x_1, x_2, x_3) = 2x_1^2 + 2x_2^2 + x_3^2 + 3x_1x_2 + x_2x_3 - 6x_1 - 4x_2 - 3x_3$$

admits a global minimum on $\mathbb{R}^3$.

The Hessian matrix is $Q = \begin{pmatrix} 4 & 3 & 0 \\ 3 & 4 & 1 \\ 0 & 1 & 2 \end{pmatrix}$

By the Matlab command $eig(Q)$ we obtain the eigenvalues of Q

$$eig(Q) \approx [0.61, 2.28, 7.09]$$

Then f is strongly convex and the global minimum point is

$x^* = -Q^{-1}c = -inv(Q) * c$, where $c = [-6, -4, -3]'$

$$x^* = -inv(Q) * c = inv(Q) * [-6, -4, -3]'$$

$x^* =$

```
 2.7000
-1.6000
 2.3000
```

Example

Check if the function

$$f(x_1, x_2, x_3) = x_1^2 + 3x_2^2 + x_3^2 + x_1x_2 - 2x_1x_3 - x_2x_3 + x_1 - x_3$$

admits a global minimum on $\mathbb{R}^3$.

The Hessian matrix is $Q = \begin{pmatrix} 2 & 1 & -2 \\ 1 & 6 & -1 \\ -2 & -1 & 2 \end{pmatrix}$, $\quad c = (1, 0, -1)^T$.

By the Matlab command $eig(Q)$ we obtain the eigenvalues of Q

$$eig(Q) \approx [0, 3.26, 6.73]$$

Then f is convex but not strongly convex and the global minimum points, if any, are given by the solutions of the system $Qx + c = 0$.

Setting $c = [1, 0, -1]'$, by the Matlab command "rank", we check that

$$rank([Q, -c]) = rank(Q) = 2,$$

which proves that the system admits solutions.

We note that the first two rows of Q are linearly independent.
Therefore, we can delete the third equation of system $Qx = -c$, which turns out to be equivalent to

$$\begin{pmatrix} 2 & 1 & -2 \\ 1 & 6 & -1 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} -1 \\ 0 \end{pmatrix}. \tag{S}$$

Setting,

$$B = \begin{pmatrix} 2 & 1 \\ 1 & 6 \end{pmatrix}, \quad N = \begin{pmatrix} -2 \\ -1 \end{pmatrix}$$

then (S) can be written as

$$B \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} + N x_3 = \begin{pmatrix} -1 \\ 0 \end{pmatrix}$$

i.e.,

$$\begin{pmatrix} 2 & 1 \\ 1 & 6 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} + \begin{pmatrix} -2 \\ -1 \end{pmatrix} x_3 = \begin{pmatrix} -1 \\ 0 \end{pmatrix}$$

Then,

$$B \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} -1 \\ 0 \end{pmatrix} - N x_3$$

and, provided that $det(B) \neq 0$, we obtain,

$$\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = B^{-1} \left[\begin{pmatrix} -1 \\ 0 \end{pmatrix} - N x_3 \right] = B^{-1} \begin{pmatrix} -1 \\ 0 \end{pmatrix} - B^{-1} N x_3$$

Computing by Matlab, $inv(B) * [-1; 0]$ and $inv(B) * N$, we obtain:

$$B^{-1} \begin{pmatrix} -1 \\ 0 \end{pmatrix} = \begin{pmatrix} -0.5455 \\ 0.0909 \end{pmatrix} \qquad B^{-1} N = \begin{pmatrix} -1 \\ 0 \end{pmatrix}$$

so that,

$$\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} -0.5455 \\ 0.0909 \end{pmatrix} - \begin{pmatrix} -1 \\ 0 \end{pmatrix} x_3 = \begin{pmatrix} -0.5455 + x_3 \\ 0.0909 \end{pmatrix}$$

The set of global minima of the function f is given by

$$X_* = \{(x_1, x_2, x_3) : x_1 = -0.5455 + x_3, x_2 = 0.0909, x_3 \in \mathbb{R}\}$$

An optimization problem $\begin{cases} \min \ f(x) \\ g(x) \leq 0 \\ h(x) = 0 \end{cases}$ is said convex if the following conditions hold:

- objective function f is convex
- inequality constraints $g_1, \ldots, g_m$ are convex functions
- equality constraints $h_1, \ldots, h_p$ are affine functions (i.e., $h_j(x) = c^\top x + d$)

Examples

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

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

but it is equivalent to the problem $\begin{cases} \min \ x_1^2 + x_2^2 \\ x_1 \leq 0 \\ x_1 + x_2 = 0 \end{cases}$ that is convex.

Main properties of convex problems

Proposition

In a convex optimization problem the feasible region X is a convex set.

Proof. The sublevel sets of convex functions are convex and the level sets of affine functions are convex. $\square$

Proposition

In a convex optimization problem any stationary point is a global optimal solution.