Constrained optimization

Mauro Passacantando

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

Numerical Methods and Optimization
Master in Computer Science – University of Pisa

Existence of global minima

A constrained optimization problem is defined as

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

where $\Omega = \{x \in \mathcal{D} : \ g(x) \leq 0, \ h(x) = 0\}$ is the feasible region.

Theorem

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 minimum.

Example.

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

admits a global minimum. Where?

Existence of global optima

Theorem

If f is continuous, 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 minimum.

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 minimum.

Existence of global optima

Theorem

If f is continuous and coercive, i.e.,

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

and Ω is closed, then there exists a global minimum.

Example.

$$\begin{cases} \min \ x^4 + 3x^3 - 5x^2 + x - 2 \\ x \geq 0 \end{cases}$$

Since f is coercive and $\Omega = \mathbb{R}_+$, there exists a global minimum.

Existence and uniqueness of global optima

Corollary

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

Example. Any quadratic programming problem

$$\begin{cases} \min \ \frac{1}{2} x^T Q x + c^T x \\ Ax \leq b \end{cases}$$

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

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

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

Theorem (Eaves)

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

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

(b) $d^{\mathsf{T}}(Qx + c) \geq 0$ for any $x \in \Omega$ and any $d \in \mathrm{rec}(\Omega)$ 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 if and only if $d^T c \geq 0 \ \forall \ d \in \text{rec}(\Omega)$
- If Q is positive definite then (a) and (b) are satisfied.
- If Ω is bounded then (a) and (b) are satisfied.

Exercise 1. Prove that the quadratic programming problem

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

has a global minimum.

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 minimum 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 minimum, then there is no descent direction in $T_\Omega(x^*)$, i.e.,

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

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. $\lim_{k\to\infty} (z_k - x^*)/t_k = d$. 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 minimum. □

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 minimum if and only if

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

Exercise 2. 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

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

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 $x \in \Omega$.

Is $T_\Omega(x) = D(x)$ true for all $x \in \Omega$? NO.

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 - Constraint qualifications

a) (*Affine constraints*)
If g_i and h_j are affine for all $i = 1, \ldots, m$ and $j = 1, \ldots, p$, then $T_\Omega(x) = D(x)$ for all $x \in \Omega$.

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

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

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

are linear independent, then $T_\Omega(\bar{x}) = D(\bar{x})$.

Karush-Kuhn-Tucker Theorem

Why condition $T_\Omega(x) = D(x)$ is important?

Theorem (Karush-Kuhn-Tucker)

If x^* is a local minimum and $T_\Omega(x^*) = D(x^*)$, then there exist $\lambda^* \in \mathbb{R}^m$ and $\mu^* \in \mathbb{R}^p$ s.t. (x^*, λ^*, μ^*) satisfies the KKT system:

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

Exercise 3. 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

Assumption $T_\Omega(x^*) = D(x^*)$ is crucial.

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 minimum. $T_\Omega(x^*) \neq D(x^*)$.

$\nabla f(x^*) = (1, 1)$, $\nabla g_1(x^*) = (0, -2)$, $\nabla g_2(x^*) = (0, 1)$.

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 minimum.

KKT Theorem for convex problems

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

Exercise 4. Prove the latter result.

Karush-Kuhn-Tucker Theorem

Exercise 5. Compute the distance between a point $z \in \mathbb{R}^n$ and the hyperplane $H = \{x \in \mathbb{R}^n : \ a^T x = b\}$

Exercise 6. Compute the distance between two parallel hyperplanes

$$H_1 = \{x \in \mathbb{R}^n : \ a^T x = b_1\}, \qquad H_2 = \{x \in \mathbb{R}^n : \ a^T x = b_2\}, \quad \text{with } b_1 \neq b_2.$$

Exercise 7. Compute the projection of a point $z \in \mathbb{R}^n$ on the ball with center x^0 and radius r.

Exercise 8. 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 minimum?

Definition (Critical cone)

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

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

Theorem

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

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 minimum, then

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

Special case: unconstrained problems

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

Second order necessary optimality condition

The previous theorem does not give a sufficient optimality condition.

Example.

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

$x^* = (0,0)$, $\lambda^* = 1$ 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 minimum 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

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

then x^* is a local minimum.

Special case: unconstrained problems.

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

Second order optimality conditions

Exercise 9. Find local and global minima of the following problems:

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

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

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

Lagrangian relaxation

Consider the general optimization problem

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

where $x \in \mathcal{D}$ and the optimal value is $v(P)$.

The Lagrangian function $L : \mathcal{D} \times \mathbb{R}^m \times \mathbb{R}^p \to \mathbb{R}$ is defined as

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

Lagrangian relaxation and dual function

Definition

Given $\lambda \geq 0$ and $\mu \in \mathbb{R}^p$, the problem

$$\begin{cases} \min \ L(x, \lambda, \mu) \\ x \in \mathcal{D} \end{cases}$$

is called Lagrangian relaxation of (P) and $\psi(\lambda, \mu) = \inf_{x \in \mathcal{D}} L(x, \lambda, \mu)$ is the Lagrangian dual function.

Dual function ψ

- is concave because inf of linear functions w.r.t (λ, μ)
- can be equal to $-\infty$
- can be not differentiable

Lagrangian relaxation and dual function

Theorem

Given $\lambda \geq 0$ and $\mu \in \mathbb{R}^p$, we have

$$\psi(\lambda, \mu) \leq v(P).$$

Proof. If $x \in \Omega$, i.e. $g(x) \leq 0, h(x) = 0$, then

$$L(x, \lambda, \mu) = f(x) + \sum_{i=1}^{m} \lambda_i \, g_i(x) \leq f(x),$$

hence

$$\psi(\lambda, \mu) = \min_{x \in \mathcal{D}} \; L(x, \lambda, \mu) \leq \min_{x \in \Omega} \; L(x, \lambda, \mu) \leq \min_{x \in \Omega} \; f(x) = v(P)$$

Lagrangian dual problem

The problem

$$\begin{cases} \max \ \psi(\lambda, \mu) \\ \lambda \geq 0 \end{cases} \tag{D}$$

is called Lagrangian dual problem of (P).

- Dual problem consists in finding the best lower bound of $v(P)$.
- Dual problem is a convex problem, even if (P) is not convex.

Lagrangian dual problem

Example - Linear Programming.

Primal problem:

$$\begin{cases} \min \ c^{\mathsf{T}} x \\ Ax \geq b \end{cases} \tag{P}$$

Lagrangian function: $L(x, \lambda) = c^{\mathsf{T}} x + \lambda^{\mathsf{T}}(b - Ax) = \lambda^{\mathsf{T}} b + (c^{\mathsf{T}} - \lambda^{\mathsf{T}} A)x$
Daul function:

$$\psi(\lambda) = \min_{x \in \mathbb{R}^n} L(x, \lambda) = \begin{cases} -\infty & \text{if } c^{\mathsf{T}} - \lambda^{\mathsf{T}} A \neq 0 \\ \lambda^{\mathsf{T}} b & \text{if } c^{\mathsf{T}} - \lambda^{\mathsf{T}} A = 0 \end{cases}$$

Dual problem:

$$\begin{cases} \max \ \psi(\lambda) \\ \lambda \geq 0 \end{cases} \longrightarrow \begin{cases} \max \ \lambda^{\mathsf{T}} b \\ \lambda^{\mathsf{T}} A = c^{\mathsf{T}} \\ \lambda \geq 0 \end{cases} \tag{D}$$

is a linear programming problem.

Exercise 10. What is the dual of (D)?

Lagrangian dual problem

Example - Least norm solution of linear equations.

Primal problem:

$$\begin{cases} \min \ x^\mathsf{T} x \\ Ax = b \end{cases} \tag{P}$$

Lagrangian function: $L(x, \mu) = x^\mathsf{T} x + \mu^\mathsf{T}(Ax - b)$.
Dual function: $\psi(\mu) = \min_{x \in \mathbb{R}^n} L(x, \mu)$.
$L(x, \mu)$ is quadratic and strongly convex w.r.t x, thus

$$\nabla_x L = 2x + A^\mathsf{T}\mu = 0 \Longleftrightarrow x = -\frac{1}{2}A^\mathsf{T}\mu,$$

hence $\psi(\mu) = -\frac{1}{4}\mu^\mathsf{T} AA^\mathsf{T}\mu - b^\mathsf{T}\mu$.
Dual problem:

$$\begin{cases} \max \ -\frac{1}{4}\mu^\mathsf{T} AA^\mathsf{T}\mu - b^\mathsf{T}\mu \\ \mu \in \mathbb{R}^p \end{cases} \tag{D}$$

is an unconstrained convex quadratic programming problem.

Lagrangian dual problem

Exercise 11. Find the dual problem of a generic quadratic programming problem

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

where Q is a symmetric positive definite matrix.

Weak duality

Theorem (weak duality)

For any optimization problem we have $v(D) \leq v(P)$.

Strong duality, i.e., $v(D) = v(P)$, does not hold in general.
Example.

$$\begin{cases} \min \ -x^2 \\ x - 1 \leq 0 \\ -x \leq 0 \end{cases} \qquad v(P) = -1$$

$L(x, \lambda) = -x^2 + \lambda_1(x - 1) - \lambda_2 x,$

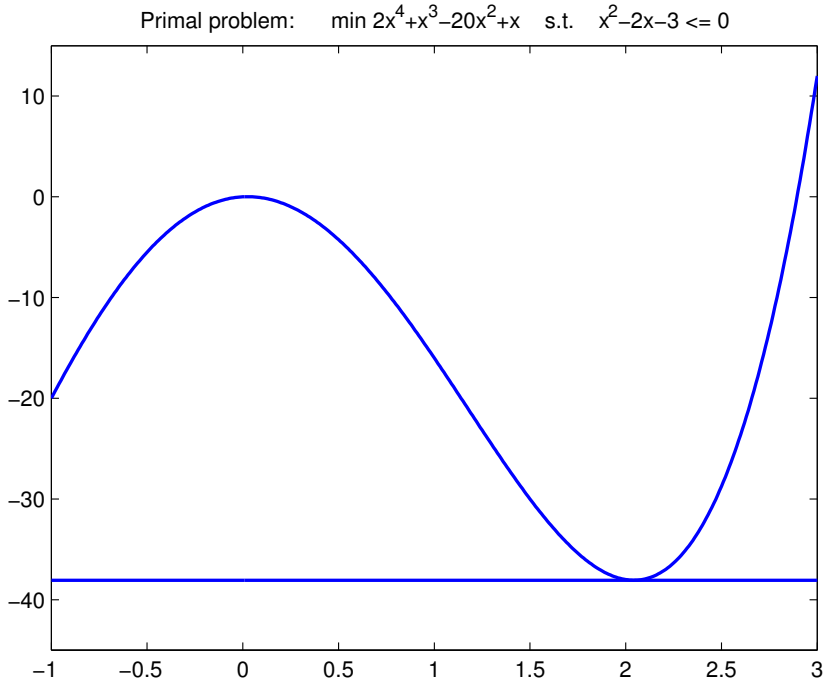
$$\psi(\lambda) = \min_{x \in \mathbb{R}} L(x, \lambda) = -\infty \qquad \forall \ \lambda \in \mathbb{R}^2,$$

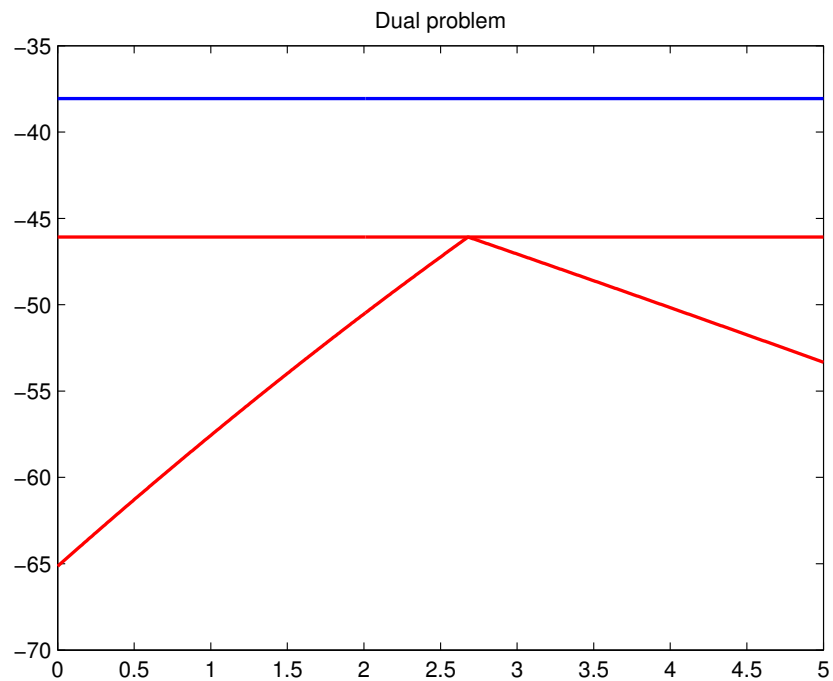
hence $v(D) = -\infty$.

Weak duality

Example.

$$\begin{cases} \min\ 2x^4 + x^3 - 20x^2 + x \\ x^2 - 2x - 3 \leq 0 \end{cases}$$





Primal optimal solution $x^* \simeq 2.0427$, $v(P) \simeq -38.0648$.
Dual optimal solution $\lambda^* \simeq 2.68$, $v(D) \simeq -46.0838$.

Strong duality

Theorem (strong duality)

If the problem

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

is convex, there exists an optimal solution x^*, and $T_\Omega(x^*) = D(x^*)$, then KKT multipliers (λ^*, μ^*) associated to x^* are an optimal solution of the dual problem and $v(D) = v(P)$.

Proof. $L(x, \lambda, \mu)$ is convex with respect to x, thus

$$v(D) \geq \psi(\lambda^*, \mu^*) = \min_x L(x, \lambda^*, \mu^*) = L(x^*, \lambda^*, \mu^*) = f(x^*) = v(P) \geq v(D)$$

□

Strong duality

Strong duality can hold also for some nonconvex problems.

Example.

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

$$L(x, \lambda) = -x_1^2 - x_2^2 + \lambda(x_1^2 + x_2^2 - 1) = (\lambda - 1)x_1^2 + (\lambda - 1)x_2^2 - \lambda.$$

$$\psi(\lambda) = \begin{cases} -\infty & \text{if } \lambda < 1 \\ -\lambda & \text{if } \lambda \geq 1 \end{cases}$$

hence $\lambda^* = 1$ is the dual optimum and $v(D) = -1$.

Exercises

Exercise 12. Consider the problem

$$\begin{cases} \min \sum_{i=1}^{n} x_i^2 \\ \sum_{i=1}^{n} x_i \geq 1 \end{cases}$$

- Discuss existence and uniqueness of optimal solutions
- Find the optimal solution and the optimal value
- Write the dual problem
- Solve the dual problem and check whether strong duality holds

Exercise 13. Given $a, b \in \mathbb{R}$ with $a < b$, consider the problem

$$\begin{cases} \min x^2 \\ a \leq x \leq b \end{cases}$$

- Find the optimal solution and the optimal value for any a, b
- Solve the dual problem and check whether strong duality holds

Equality constrained problems

Consider

$$\begin{cases} \min \ f(x) \\ Ax = b \end{cases}$$

with

- f strongly convex and twice continuously differentiable
- A matrix $p \times n$ with $\text{rank}(A) = p$

It is equivalent to an unconstrained problem:

write $A = (A_B, A_N)$ with $det(A_B) \neq 0$, then $Ax = b$ is equivalent to

$$A_B x_B + A_N x_N = b \quad \Longrightarrow \quad x_B = A_B^{-1}(b - A_N x_N),$$

thus

$$\begin{cases} \min \ f(x) \\ Ax = b \end{cases} \quad \text{is equivalent to} \quad \begin{cases} \min \ f(A_B^{-1}(b - A_N x_N), x_N) \\ x_N \in \mathbb{R}^{n-p} \end{cases}$$

Equality constrained problems

Example. Consider

$$\begin{cases} \min\ x_1^2 + x_2^2 + x_3^2 \\ x_1 + x_3 = 1 \\ x_1 + x_2 - x_3 = 2 \end{cases}$$

Since $x_1 = 1 - x_3$ and $x_2 = 2 - x_1 + x_3 = 1 + 2x_3$, the original constrained problem is equivalent to the following unconstrained problem:

$$\begin{cases} \min\ (1 - x_3)^2 + (1 + 2x_3)^2 + x_3^2 = 6x_3^2 + 2x_3 + 2 \\ x_3 \in \mathbb{R} \end{cases}$$

Therefore, the optimal solution is $x_3 = -1/6$, $x_1 = 7/6$, $x_2 = 2/3$.

Frank-Wolfe method

Consider the problem

$$\begin{cases} \min \ f(x) \\ Ax \leq b \end{cases} \tag{P}$$

where f is convex and continuously differentiable and the polyhedron $\Omega = \{x : \ Ax \leq b\}$ is bounded.

Linearize the objective function at x^k, i.e., $f(x) \simeq f(x^k) + \nabla f(x^k)^{\mathsf{T}}(x - x^k)$ and solve the linear programming problem:

$$\begin{cases} \min \ f(x^k) + \nabla f(x^k)^{\mathsf{T}}(x - x^k) \\ A\,x \leq b \end{cases}$$

The optimal solution y^k of the linearized problem gives a lower bound for (P):

$$f(x^k) \geq v(P) = \min_{x \in \Omega} f(x) \geq \min_{x \in \Omega} \left[f(x^k) + \nabla f(x^k)^{\mathsf{T}}(x - x^k) \right] = f(x^k) + \nabla f(x^k)^{\mathsf{T}}(y^k - x^k)$$

If $\nabla f(x^k)^{\mathsf{T}}(y^k - x^k) = 0$ then x^k solves (P),
otherwise $\nabla f(x^k)^{\mathsf{T}}(y^k - x^k) < 0$, i.e., $y^k - x^k$ is a descent direction for f at x^k.

Frank-Wolfe method

Frank-Wolfe method

0. Choose a feasible point x_0 and set $k = 0$

1. Compute an optimal solution y^k of the LP problem

$$\begin{cases} \min \ \nabla f(x^k)^{\mathsf{T}} x \\ Ax \leq b \end{cases}$$

2. **if** $\nabla f(x^k)^{\mathsf{T}}(y^k - x^k) = 0$ **then** STOP
else compute step size t_k

3. Set $x^{k+1} = x^k + t_k(y^k - x^k)$, $k = k + 1$ and go to step 1.

The step size t_k can be predetermined, e.g., $t_k = \dfrac{1}{k+1}$
or computed by an (exact/inexact) line search.

Theorem

For any starting point x_0 the generated sequence $\{x^k\}$ is bounded and any of its cluster points is a global minimum.

Frank-Wolfe method

Example. Solve the problem

$$\begin{cases} \min\ (x_1-3)^2+(x_2-1)^2 \\ 0 \le x_1 \le 2 \\ 0 \le x_2 \le 2 \end{cases}$$

by means of the Frank-Wolfe method starting from the point $x^0=(0,0)$ and using an exact line search to compute the step size.

$\nabla f(x)=(2x_1-6,\ 2x_2-2)$ and $\nabla f(x^0)=(-6,\ -2)$, the optimal solution of the linearized problem

$$\begin{cases} \min\ -6x_1-2x_2 \\ x \in \Omega \end{cases}$$

is $y^0=(2,2)$. Since $\nabla f(x^0)^{\mathsf{T}}(y^0-x^0)=-56$, a line search is needed: the step size t_0 is the optimal solution of

$$\begin{cases} \min\ (x_1-3)^2+(x_2-1)^2 \\ x_1=2t \\ x_2=2t \\ t \in [0,1] \end{cases}$$

Frank-Wolfe method

which is equivalent to

$$\begin{cases} \min\ 8t^2 - 16t \\ t \in [0,1] \end{cases}$$

thus $t_0 = 1$ and $x^1 = (2,2)$. Since $\nabla f(x^1) = (-2,2)$, the optimal solution of the linearized problem

$$\begin{cases} \min\ -2x_1 + 2x_2 \\ x \in \Omega \end{cases}$$

is $y^1 = (2,0)$. Now, $\nabla f(x^1)^{\mathsf{T}}(y^1 - x^1) = -4$, the step size is the optimal solution of

$$\begin{cases} \min\ (x_1 - 3)^2 + (x_2 - 1)^2 \\ x_1 = 2 \\ x_2 = 2 - 2t \\ t \in [0,1] \end{cases}$$

hence $t_1 = 1/2$, $x^2 = (2,1)$ and $\nabla f(x^2) = (-2,0)$, thus $y^2 = (2,0)$ (an alternative optimal solution is $y^2 = (2,2)$). Since $\nabla f(x^2)^{\mathsf{T}}(y^2 - x^2) = 0$, the point x^2 is the optimal solution of the original problem.

Frank-Wolfe method

Exercise 14. Implement in MATLAB the Frank-Wolfe method with exact line search for solving the problem

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

where Q is a positive definite matrix.

Exercise 15. Run the Frank-Wolfe method with exact line search for solving the problem

$$\begin{cases} \min \ \frac{1}{2}(x_1 - 3)^2 + (x_2 - 2)^2 \\ -2x_1 + x_2 \leq 0 \\ x_1 + x_2 \leq 4 \\ -x_2 \leq 0 \end{cases}$$

starting from the point $(0, 0)$.
[Use $\nabla f(x^k)^{\mathsf{T}}(x^k - y^k) < 10^{-3}$ as stopping criterion.]

Exercise 16. Solve the previous problem by means of the Frank-Wolfe method with predetermined step size $t_k = 1/(k+1)$ starting from $(0, 0)$.

Penalty method

Consider a constrained optimization problem

$$\begin{cases} \min \ f(x) \\ g_i(x) \leq 0 \qquad \forall \ i = 1, \dots, m \end{cases} \tag{P}$$

Define the penalty function

$$p(x) = \sum_{i=1}^{m} \left(\max\{0, g_i(x)\}\right)^2$$

and consider the (unconstrained) penalized problem

$$\begin{cases} \min \ f(x) + \dfrac{1}{\varepsilon} p(x) := p_\varepsilon(x) \\ x \in \mathbb{R}^n \end{cases} \tag{P_ε}$$

Note that

$$p_\varepsilon(x) \begin{cases} = f(x) & \text{if } x \in \Omega \\ > f(x) & \text{if } x \notin \Omega \end{cases}$$

Penalty method

Proposition.

- If f, g_i are continuously differentiable, then p_ε is continuously differentiable
- If f, g_i are convex, then p_ε is convex
- (P_ε) is a relaxation of (P), i.e., $v(P_\varepsilon) \leq v(P)$ for any $\varepsilon > 0$
- Let x_ε^* be an optimal solution of (P_ε):
  - if $x_\varepsilon^* \in \Omega$, then x_ε^* is optimal also for (P)
  - if $x_\varepsilon^* \notin \Omega$, then $v(P_\varepsilon) > v(P_{\varepsilon'})$ for any $\varepsilon' > \varepsilon$

Penalty method

0. Set $\varepsilon_0 > 0$, $\tau \in (0,1)$, $k = 0$

1. Find an optimal solution x^k of (P_{ε_k})

2. **If** $x^k \in \Omega$ **then** STOP
else $\varepsilon_{k+1} = \tau\varepsilon_k$, $k = k+1$ and go to step 1.

Theorem

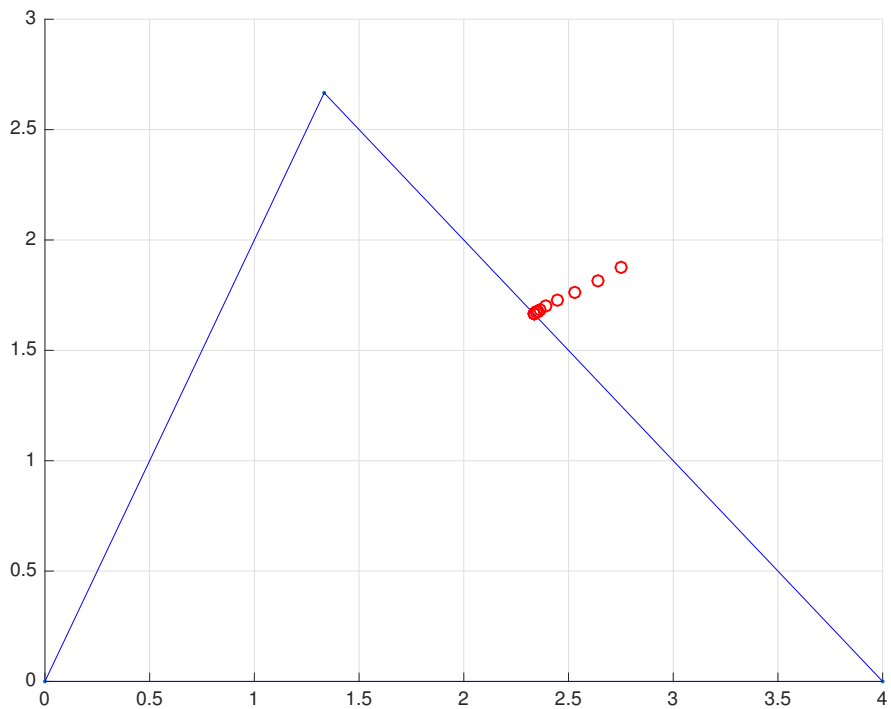
If f is coercive, then the sequence $\{x^k\}$ is bounded and any of its cluster points is an optimal solution of (P).

Penalty method

Example. Solve the problem

$$\begin{cases} \min \ \frac{1}{2}(x_1 - 3)^2 + (x_2 - 2)^2 \\ -2x_1 + x_2 \leq 0 \\ x_1 + x_2 \leq 4 \\ -x_2 \leq 0 \end{cases}$$

by means of the penalty method starting from $\varepsilon_0 = 5$ with $\tau = 0.5$.



Penalty method

Exercise 17. Implement in MATLAB the penalty method for solving the problem

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

where Q is a positive definite matrix.

Exercise 18. Run the penalty method with $\tau = 0.5$ and $\varepsilon_0 = 5$ for solving the problem

$$\begin{cases} \min \ \frac{1}{2}(x_1 - 3)^2 + (x_2 - 2)^2 \\ -2x_1 + x_2 \leq 0 \\ x_1 + x_2 \leq 4 \\ -x_2 \leq 0 \end{cases}$$

[Use $\min(b - Ax) > -10^{-3}$ as stopping criterion.]

Barrier methods

Consider

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

where

- f, g_i convex and twice continuously differentiable
- there is no isolated point in Ω
- there exists an optimal solution (e.g. f coercive or $\{g(x) \leq 0\}$ bounded)
- Slater constraint qualification holds: there exists $\bar{x}$ such that

$$\bar{x} \in \mathrm{dom(f)}, \qquad g_i(\bar{x}) < 0, \ i = 1, \ldots, m$$

Hence strong duality holds.

Special cases: linear programming, convex quadratic programming

Unconstrained reformulation

The constrained problem

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

is equivalent to the unconstrained problem

$$\begin{cases} \min \ f(x) + \sum_{i=1}^{m} I_-(g_i(x)) \\ x \in \mathbb{R}^n \end{cases}$$

where

$$I_-(u) = \begin{cases} 0 & \text{if } u \leq 0 \\ +\infty & \text{if } u > 0 \end{cases}$$

is the indicator function of $\mathbb{R}_-$.

Barrier function

I_- is neither finite nor differentiable. It can be approximated by the smooth convex function

$$u \mapsto -\frac{\varepsilon}{u}, \qquad \text{where } u < 0,$$

and parameter $\varepsilon > 0$. Approximation improved as $\varepsilon \to 0$.
We approximate the problem

$$\begin{cases} \min \ f(x) + \sum_{i=1}^{m} I_-(g_i(x)) \\ x \in \mathbb{R}^n \end{cases}$$

with

$$\begin{cases} \min \ f(x) + \varepsilon \sum_{i=1}^{m} \frac{1}{-g_i(x)} \\ x \in \text{int}(\Omega) \end{cases}$$

$B(x) = \sum_{i=1}^{m} \frac{1}{-g_i(x)}$ is called barrier function.
$\text{dom}(B) = \text{int}(\Omega)$, B is convex and smooth.

Barrier method

If $x^*(\varepsilon)$ is the optimal solution of

$$\begin{cases} \min \ f(x) + \varepsilon \sum_{i=1}^{m} \dfrac{1}{-g_i(x)} \\ x \in \text{int}(\Omega) \end{cases}$$

then

$$\nabla f(x^*(\varepsilon)) + \varepsilon \sum_{i=1}^{m} \frac{1}{[g_i(x^*(\varepsilon))]^2} \nabla g_i(x^*(\varepsilon)) = 0.$$

Define $\lambda_i^*(\varepsilon) = \dfrac{\varepsilon}{[g_i(x^*(\varepsilon))]^2} > 0$, for any $i = 1, \ldots, m$. Then the Lagrangian function

$$L(x, \lambda^*(\varepsilon)) = f(x) + \sum_{i=1}^{m} \lambda_i^*(\varepsilon) g_i(x)$$

is convex and $\nabla_x L(x^*(\varepsilon), \lambda^*(\varepsilon)) = 0$, hence

$$\begin{aligned} f(x^*(\varepsilon)) \geq v(P) \geq \psi(\lambda^*(\varepsilon)) \quad &= \min_x L(x, \lambda^*(\varepsilon)) \\ &= L(x^*(\varepsilon), \lambda^*(\varepsilon)) \\ &= f(x^*(\varepsilon)) - \varepsilon B(x^*(\varepsilon)) \end{aligned}$$

Barrier method

Barrier method

0. Set $\tau < 1$ and $\varepsilon_1 > 0$. Choose $x^0 \in \text{int}(\Omega)$, set $k = 1$

1. Find the optimal solution x^k of

$$\begin{cases} \min \ f(x) + \varepsilon_k \displaystyle\sum_{i=1}^{m} \frac{1}{-g_i(x)} \\ x \in \text{int}(\Omega) \end{cases}$$

using x^{k-1} as starting point

2. **If** $\varepsilon_k B(x^k) = 0$ **then** STOP
else $\varepsilon_{k+1} = \tau \varepsilon_k$, $k = k + 1$ and go to step 1

Theorem

If Ω is bounded, then the sequence $\{x^k\}$ is bounded and any of its cluster points is an optimal solution of (P).

Logarithmic barrier

The indicator function I_- can be also approximated by the smooth convex function:

$$u \mapsto -\varepsilon \log(-u), \qquad \text{with } \varepsilon > 0.$$

We approximate the problem

$$\begin{cases} \min \ f(x) + \sum_{i=1}^{m} I_-(g_i(x)) \\ x \in \mathbb{R}^n \end{cases}$$

with

$$\begin{cases} \min \ f(x) - \varepsilon \sum_{i=1}^{m} log(-g_i(x)) \\ x \in \text{int}(\Omega) \end{cases}$$

Logarithmic barrier

Logarithmic barrier function

$$B(x) = -\sum_{i=1}^{m} log(-g_i(x))$$

- $\mathrm{dom}(B) = \mathrm{int}(\Omega)$.
- B is convex
- B is smooth with

$$\nabla B(x) = -\sum_{i=1}^{m} \frac{1}{g_i(x)} \nabla g_i(x)$$

$$\nabla^2 B(x) = \sum_{i=1}^{m} \frac{1}{g_i(x)^2} \nabla g_i(x) \nabla g_i(x)^\mathsf{T} + \sum_{i=1}^{m} \frac{1}{-g_i(x)} \nabla^2 g_i(x)$$

Logarithmic barrier

If $x^*(\varepsilon)$ is the optimal solution of

$$\begin{cases} \min \ f(x) - \varepsilon \sum_{i=1}^{m} log(-g_i(x)) \\ x \in \mathbb{R}^n \end{cases}$$

then

$$\nabla f(x^*(\varepsilon)) + \sum_{i=1}^{m} \frac{\varepsilon}{-g_i(x^*(\varepsilon))} \nabla g_i(x^*(\varepsilon)) = 0.$$

Define $\lambda_i^*(\varepsilon) = \dfrac{\varepsilon}{-g_i(x^*(\varepsilon))} > 0$, for any $i = 1, \ldots, m$. Then the Lagrangian function

$$L(x, \lambda^*(\varepsilon)) = f(x) + \sum_{i=1}^{m} \lambda_i^*(\varepsilon) g_i(x)$$

is convex and $\nabla_x L(x^*(\varepsilon), \lambda^*(\varepsilon)) = 0$, hence

$$\begin{aligned} f(x^*(\varepsilon)) \geq v(P) \geq \psi(\lambda^*(\varepsilon)) \quad &= \min_x L(x, \lambda^*(\varepsilon)) \\ &= L(x^*(\varepsilon), \lambda^*(\varepsilon)) \\ &= f(x^*(\varepsilon)) - m\varepsilon \end{aligned}$$

Interpretation via KKT conditions

KKT system of the original problem is

$$\begin{cases} \nabla f(x) + \sum_{i=1}^{m} \lambda_i \nabla g_i(x) = 0 \\ -\lambda_i g_i(x) = 0 \\ \lambda \geq 0 \\ g(x) \leq 0 \end{cases}$$

$(x^*(\varepsilon), \lambda^*(\varepsilon))$ solves the system

$$\begin{cases} \nabla f(x) + \sum_{i=1}^{m} \lambda_i \nabla g_i(x) = 0 \\ -\lambda_i g_i(x) = \varepsilon \\ \lambda \geq 0 \\ g(x) \leq 0 \end{cases}$$

which is an approximation of the KKT system

Logarithmic barrier method

Logarithmic barrier method

0. Set tolerance $\delta > 0$, $\tau < 1$ and $\varepsilon_1 > 0$. Choose $x^0 \in \text{int}(\Omega)$, set $k = 1$

1. Find the optimal solution x^k of

$$\begin{cases} \min \; f(x) - \varepsilon_k \sum_{i=1}^{m} log(-g_i(x)) \\ x \in \text{int}(\Omega) \end{cases}$$

using x^{k-1} as starting point

2. **If** $m\,\varepsilon_k < \delta$ **then** STOP
else $\varepsilon_{k+1} = \tau \varepsilon_k$, $k = k + 1$ and go to step 1

Choice of τ involves a trade-off: small τ means fewer outer iterations, more inner iterations

Choice of starting point

How to find $x^0 \in \text{int}(\Omega)$?

Consider the auxiliary problem

$$\begin{cases} \min \ s \\ g_i(x) \leq s \end{cases}$$

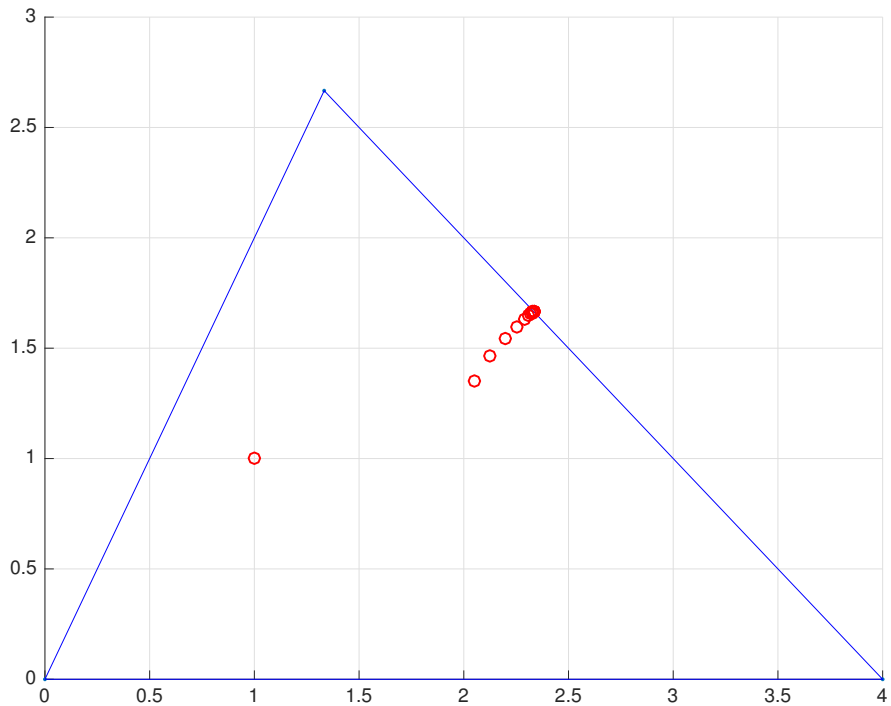
- Take any $\tilde{x} \in \mathbb{R}^n$ and $\tilde{s} > \max_i g_i(\tilde{x})$ so that $(\tilde{x}, \tilde{s})$ is in the interior of the feasible region of the auxiliary problem
- Find an optimal solution (x^*, s^*) of the auxiliary problem with the barrier method starting from $(\tilde{x}, \tilde{s})$
- If $s^* < 0$, then $x^* \in \text{int}(\Omega)$
  otherwise $\text{int}(\Omega) = \emptyset$

Logarithmic barrier method

Example. Solve the problem

$$\begin{cases} \min \ \frac{1}{2}(x_1 - 3)^2 + (x_2 - 2)^2 \\ -2x_1 + x_2 \leq 0 \\ x_1 + x_2 \leq 4 \\ -x_2 \leq 0 \end{cases}$$

by means of the logarithmic barrier method with $\delta = 10^{-3}$, $\tau = 0.5$, $\varepsilon_1 = 1$ and $x^0 = (1, 1)$.



Logarithmic barrier method

Exercise 19. Implement in MATLAB the logarithmic barrier method for solving the problem

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

where Q is a positive definite matrix.

Exercise 20. Run the logarithmic barrier method with $\delta = 10^{-3}$, $\tau = 0.5$, $\varepsilon_1 = 1$ and $x^0 = (1, 1)$ for solving the problem

$$\begin{cases} \min \ \frac{1}{2}(x_1 - 3)^2 + (x_2 - 2)^2 \\ -2x_1 + x_2 \leq 0 \\ x_1 + x_2 \leq 4 \\ -x_2 \leq 0 \end{cases}$$