8 - Solution methods for constrained optimization problems

Mauro Passacantando

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

Optimization Methods and Game Theory
Master of Science in Artificial Intelligence and Data Engineering
University of Pisa – A.Y. 2020/21

Problems with linear equality constraints

Consider a constrained problem

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

where

- f is strongly convex and twice continuously differentiable
- A is $p \times n$ matrix 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

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

Problems with linear equality constraints

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 $x_3 = -1/6$, $x_1 = 7/6$, $x_2 = 2/3$.

Active-set method

Consider a 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 positive definite
- for any feasible point x the vectors $\{A_i : \ A_i x = b_i\}$ are linearly independent

The active-set method solves at each iteration a quadratic programming problem with equality constraints only.

Active-set method

0. Choose a feasible point x^0, set $W_0 = \{i : \; A_i x^0 = b_i\}$ (working set) and $k = 0$.

1. Find the optimal solution y^k of the problem

$$\begin{cases} \min \; \frac{1}{2} x^\mathsf{T} Q x + c^\mathsf{T} x \\ A_i x = b_i \qquad \forall \; i \in W_k \end{cases}$$

2. **If** $y^k \neq x^k$ **then** go to step 3
else go to step 4

3. **If** y^k is feasible **then** $t_k = 1$
else $t_k = \min \left\{ \dfrac{b_i - A_i x^k}{A_i(y^k - x^k)} : \; i \notin W_k, \; A_i(y^k - x^k) > 0 \right\}$,
end
$x^{k+1} = x^k + t_k(y^k - x^k)$, $W_{k+1} = W_k \cup \{i \notin W_k : \; A_i x^{k+1} = b_i\}$,
$k = k + 1$ and go to step 1

4. Compute the KKT multipliers μ^k related to y^k
If $\mu^k \geq 0$ **then** STOP
else $x^{k+1} = x^k$, $\mu^k_j = \min\limits_{i \in W_k} \mu^k_i$, $W_{k+1} = W_k \setminus \{j\}$, $k = k + 1$ and go to step 1

Active-set method

Example. Solve the problem

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

by means of the active-set method starting from $x^0=(0,0)$.

The working set $W_0=\{1,3\}$ hence $y^0=x^0$; KKT multipliers are $\mu_1^0=-3/2$, $\mu_3^0=-11/2$. The new point $x^1=x^0$ with $W_1=\{1\}$; y^1 is the optimal solution of

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

which is equivalent to

$$\begin{cases} \min \ \frac{9}{2}x_1^2-11x_1 \\ x_1 \in \mathbb{R} \end{cases}$$

thus $y^1=(11/9,22/9)$ which is feasible, therefore $x^2=y^1$ and $W_2=\{1\}$. We already know that $y^2=x^2$; the KKT multiplier is $\mu_1^2=-8/9$, hence $x^3=x^2$ and $W_3=\emptyset$.

Active-set method

The optimal solution $y^3 = (3, 2)$ is not feasible, the step size is

$$t_3 = \min\left\{ \frac{b_2 - A_2 x^3}{A_2(y^3 - x^3)} \;,\; \frac{b_3 - A_3 x^3}{A_3(y^3 - x^3)} \right\} = \min\left\{ \frac{1}{4}, \frac{11}{2} \right\} = \frac{1}{4},$$

$x^4 = x^3 + t_3(y^3 - x^3) = (5/3, 7/3)$ and $W_4 = \{2\}$. The optimal solution $y^4 = (7/3, 5/3)$ is feasible, hence $x^5 = y^4$ and $W_5 = \{2\}$. Finally, $y^5 = x^5$ and $\mu_2^5 = 2/3 > 0$, thus x^5 is the global minimum of the original problem.

Active-set method

Exercise 8.1. Solve the problem

$$\begin{cases} \min \ 2x_1^2 + x_2^2 - x_1 x_2 - 2x_1 + x_2 \\ -x_1 \leq 0 \\ -x_2 \leq 0 \\ x_1 + x_2 \leq 12 \end{cases}$$

by means of the active-set method starting from the point $(0, 10)$.

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 quadratic penalty function

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

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 and $\nabla p_\varepsilon(x) = \nabla f(x) + \dfrac{2}{\varepsilon} \sum_{i=1}^{m} \max\{0, g_i(x)\} \nabla g_i(x)$
- If f and g_i are convex, then p_ε is convex
- Any (P_ε) is a relaxation of (P), i.e., $v(P_\varepsilon) \leq v(P)$ for any $\varepsilon > 0$
- If $0 < \varepsilon_1 < \varepsilon_2$, then $v(P_{\varepsilon_2}) \leq v(P_{\varepsilon_1})$
- If x^*_ε solves (P_ε) and $x^*_\varepsilon \in \Omega$, then x^*_ε is optimal also for (P)

Penalty method

Proposition

- If f, g_i are continuously differentiable, then p_ε is continuously differentiable and $\nabla p_\varepsilon(x) = \nabla f(x) + \dfrac{2}{\varepsilon} \sum_{i=1}^{m} \max\{0, g_i(x)\} \nabla g_i(x)$
- If f and g_i are convex, then p_ε is convex
- Any (P_ε) is a relaxation of (P), i.e., $v(P_\varepsilon) \leq v(P)$ for any $\varepsilon > 0$
- If $0 < \varepsilon_1 < \varepsilon_2$, then $v(P_{\varepsilon_2}) \leq v(P_{\varepsilon_1})$
- If x_ε^* solves (P_ε) and $x_\varepsilon^* \in \Omega$, then x_ε^* is optimal also for (P)

Penalty method

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

1. Find an optimal solution x^k of the penalized problem (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.

Penalty method

Theorem

- If f is coercive, then the sequence $\{x^k\}$ is bounded and any of its cluster points is an optimal solution of (P).
- If $\{x^k\}$ converges to x^*, then x^* is an optimal solution of (P).
- If $\{x^k\}$ converges to x^* and the gradients of active constraints at x^* are linear independent, then x^* is an optimal solution of (P) and the sequence of vectors $\{\lambda^k\}$ defined as

$$\lambda_i^k := \frac{2}{\varepsilon_k} \max\{0, g_i(x^k)\}, \qquad i = 1, \ldots, m$$

converges to a vector λ^* of KKT multipliers associated to x^*.

Penalty method

Exercise 8.2.

a) Implement in MATLAB the penalty method for solving the problem

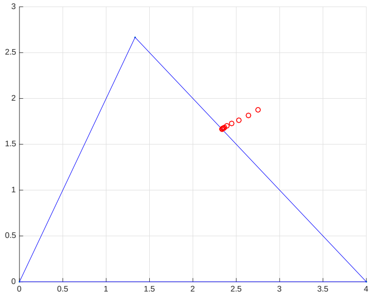
$$\begin{cases} \min \ \frac{1}{2}x^\mathsf{T}Qx + c^\mathsf{T}x \\ Ax \le b \end{cases}$$

where Q is a positive definite matrix.

b) Run the penalty method with $\tau = 0.5$ and $\varepsilon_0 = 5$ for solving the problem

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

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



Exact penalty method

Consider a convex constrained problem

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

and define the linear penalty function

$$\widetilde{p}(x) = \sum_{i=1}^{m} \max\{0, g_i(x)\}.$$

Then the penalized problem

$$\begin{cases} \min \ f(x) + \dfrac{1}{\varepsilon}\, \widetilde{p}(x) \\ x \in \mathbb{R}^n \end{cases} \tag{$\widetilde{P}_\varepsilon$}$$

is unconstrained, convex and nonsmooth.
However, we do not need a sequence $\varepsilon_k \to 0$ to approximate an optimal solution of (P) ($\to$ avoid numerical issues).

Exact penalty method

Proposition

Suppose that there exists an optimal solution x^* of (P) and λ^* is a KKT multipliers vector associated to x^*. Then, the sets of optimal solutions of (P) and $(\widetilde{P}_\varepsilon)$ coincide provided that $\varepsilon \in (0, 1/\|\lambda^*\|_\infty)$.

Exact penalty method

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

1. Find an optimal solution x^k of the penalized problem $(\widetilde{P}_{\varepsilon_k})$

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

Theorem

The exact penalty method stops after a finite number of iterations at an optimal solution of (P).

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 Ω bounded)
- Slater constraint qualification holds: there exists $\bar{x}$ such that

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

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 called the indicator function of $\mathbb{R}_{-}$, that is neither finite nor differentiable.

Logarithmic barrier

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

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

and the approximation improves as $\varepsilon \to 0$.

Hence, we can approximate the problem

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

with

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

Remark. I_- can also be approximated by the smooth convex function

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

Another barrier method is based on this approximation.

Logarithmic barrier

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

is called logarithmic barrier function. It has the following properties:

- $\text{dom}(B) = \text{int}(\Omega)$
- 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^*_ε is the optimal solution of

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

then

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

Define $\lambda^*_\varepsilon = \left(\frac{\varepsilon}{-g_1(x^*_\varepsilon)}, \dots, \frac{\varepsilon}{-g_m(x^*_\varepsilon)} \right) > 0.$ Then the Lagrangian function

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

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

$$f(x^*_\varepsilon) \geq v(P) \geq \varphi(\lambda^*_\varepsilon) = \min_x L(x, \lambda^*_\varepsilon) \quad = L(x^*_\varepsilon, \lambda^*_\varepsilon) = f(x^*_\varepsilon) - \underbrace{m\,\varepsilon}_{\text{optimality gap}}$$

Interpretation via KKT conditions

The KKT system of the original problem is

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

Notice that $(x^*_\varepsilon, \lambda^*_\varepsilon)$ 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 above 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

$$\begin{cases} \min\limits_{x,s} \ s \\ g_i(x) \le s \end{cases}$$

- Take any $\tilde{x} \in \mathbb{R}^n$, find $\tilde{s} > \max\limits_{i=1,\dots,m} g_i(\tilde{x})$
  [$(\tilde{x}, \tilde{s})$ is in the interior of the feasible region of the auxiliary problem]
- Find an optimal solution (x^*, s^*) of the auxiliary problem using a barrier method starting from $(\tilde{x}, \tilde{s})$
- If $s^* < 0$ then $x^* \in \text{int}(\Omega)$
  else $\text{int}(\Omega) = \emptyset$

Logarithmic barrier method

Exercise 8.3.

a) Implement in MATLAB the logarithmic barrier 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.

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

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