Continuous optimization
PGE305

Laurent Pfeiffer
Inria and CentraleSupélec, Université Paris-Saclay

Ensta-Paris
Institut Polytechnique de Paris
November 9, 2021

Introduction

Aim of the lecture: a general presentation of two numerical methods for constrained optimization.

- **Penalty methods** $\rightsquigarrow$ equality constraints
- **Projected gradient methods** $\rightsquigarrow$ inequality constraints
- **Interior point methods** $\rightsquigarrow$ inequality constraints.

Warning:

- The methods are general **tools**, which needs to **specified**, improved, combined with each other... depending on the nature of the problem of interest.
- Theoretical justifications, implementation details are **skipped**.

Reference:

Nocedal and Wright. Numerical optimization. Springer Science and Business Media, 2006.

Quadratic penalization

We consider in this section

$$\inf_{x\in\mathbb{R}^n} f(x), \quad \text{subject to: } g(x) = 0, \qquad (P)$$

where $f : \mathbb{R}^n \to \mathbb{R}$ and $g : \mathbb{R}^n \to \mathbb{R}^m$ are given and "smooth".

A general difficulty: we need to cope with **two general goals**:

- Minimizing f
- Ensuring the feasibility of x.

When designing a numerical method, the question arises:
Given an iterate x_k, should we look for x_{k+1} so that

$$f(x_{k+1}) < f(x_k) \quad \text{or} \quad \|g(x_{k+1})\| < \|g(x_k)\| \quad ?$$

Quadratic penalization

Main idea: combining the two objectives into a single one.
Given a real number $c \geq 0$, consider the **penalty problem:**

$$\inf_{x \in \mathbb{R}^n} Q_c(x) := f(x) + \frac{c}{2}\|g(x)\|^2. \qquad (P(c))$$

Vocabulary:

- c is called **penalty parameter**
- the expression $\frac{c}{2}\|g(x)\|^2$ is called penalty term $\rightsquigarrow$ it replaces the constraint!

A rough statement: if c is large, (P) and (P_c) are "almost" equivalent.

Intuition. For a solution x_c to problem (P_c), both $f(x_c)$ and $\frac{c}{2}\|g(x_c)\|^2$ are expected to be small. $\rightarrow$ If c is large, then $\|g(x_c)\|$ is very small.

Quadratic penalization

Considering $(P(c))$ instead of (P), we make a **trade-off** between the two objectives stated before.

Relative importance of the two objectives is represented by c:

- c is small $\rightarrow$ minimality of f matters more
- c is large $\rightarrow$ feasibility of the constraint matters more.

Big advantage of the approach: numerical **methods of unconstrained optimization** can be employed for solving (P_c).

Quadratic penalization

Exercise.
Consider the problem:

$$\inf_{x\in\mathbb{R}} x, \quad \text{subject to: } x = 0.$$

1. What is the solution $\bar{x}$ to the problem?
2. Calculate the solution x_c to the corresponding penalized problem P_c.
3. Verify that $x_c \underset{c\to+\infty}{\longrightarrow} \bar{x}$.

Quadratic penalization

Solution.

1. Obviously $\bar{x} = 0$, since 0 is the unique feasible point of the problem.

2. Let $c > 0$. We have $Q_c(x) = x + \frac{c}{2}x^2$ and $\nabla Q_c(x) = 1 + cx$. Therefore,

$$\nabla Q_c(x) = 0 \Longleftrightarrow x = -\frac{1}{c}.$$

Since Q_c is convex, $x_c := -1/c$ is the unique solution of (P_c).

3. Obviously

$$x_c = -1/c \underset{c\to\infty}{\longrightarrow} 0 = \bar{x}.$$

Quadratic penalization

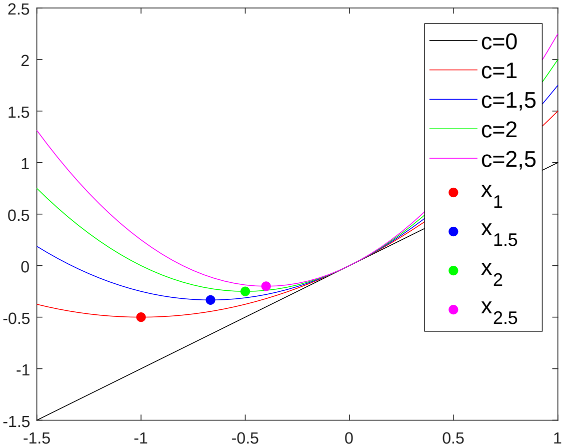


Figure: Graph of Q_c, for various values of c

Quadratic penalization

Lemma 1

Let $c_k \to \infty$. Let $(x_k)_{k\in\mathbb{N}}$ be a sequence in $\mathbb{R}^n$. Assume that

- *For all $k \in \mathbb{N}$, x_k is the* solution *to (P_{c_k}).*
- *The sequence $(x_k)_{k\in\mathbb{N}}$* converges, *let $\bar{x}$ denote the limit.*
- *There exists $\tilde{x}$ such that $g(\tilde{x}) = 0$.*

Then, $\bar{x}$ is a solution *to the original constrained problem (P).*

Proof. Step 1. Let x be a feasible point (that is, $g(x) = 0$). Then,

$$Q_{c_k}(x) = f(x) + \frac{c_k}{2}\|g(x)\|^2 = f(x).$$

In particular, $Q_{c_k}(\tilde{x}) = f(\tilde{x})$.

Quadratic penalization

Step 2: $\bar{x}$ is feasible. For all $k \in \mathbb{N}$, we have

$$\begin{aligned} c_k \|g(x_k)\|^2 &= Q_{c_k}(x_k) - f(x_k) \\ &\leq Q_{c_k}(\tilde{x}) - f(x_k) && [\textit{Optimality of } x_k] \\ &= f(\tilde{x}) - f(x_k). && [\textit{Equality of Step 1}] \end{aligned}$$

Since $f(x_k) \to f(\bar{x})$, the sequence $(f(x_k))_{k \in \mathbb{N}}$ is bounded. Therefore, there exist $M > 0$ such that $c_k \|g(x_k)\|^2 \leq M$. Thus

$$\|g(x_k)\| \leq \sqrt{M/c_k}, \quad \forall k \in \mathbb{N}.$$

Passing to the limit, we get $\|g(\bar{x})\| \leq 0$. Thus $\bar{x}$ is **feasible**.

Quadratic penalization

Step 3. Optimality of $\bar{x}$. Let x be feasible. We have

$$\begin{aligned} f(x_k) &\leq f(x_k) + c_k \|g(x_k)\|^2 \\ &= Q_{c_k}(x_k) \\ &\leq Q_{c_k}(x) && [\textit{Optimality of } x_k] \\ &= f(x). && [\textit{Equality of Step 1}] \end{aligned}$$

Passing to the limit, we get

$$f(\bar{x}) \leq f(x).$$

Thus $\bar{x}$ is optimal.

Quadratic penalization

The result of the lemma must be seen as an "ideal" situation.

Difficulties in practice:

- The problem (P_c) **may not have a solution**, even if (P) has a solution. Example:

$$\inf_{x \in \mathbb{R}} x^3, \quad \text{subject to: } x = 0.$$

- The sequence $(x_k)_{k \in \mathbb{N}}$ may not converge.
- If x_k is only a local solution of (P_{c_k}), no guaranty that $\bar{x}$ is a local solution of (P).
- The problem (P_c) is **hard to solve** when c is large, it is likely to be ill-conditioned (see next example).

Quadratic penalization

Example. Consider:

$$\inf_{(x,y)\in\mathbb{R}^2} \frac{1}{2}\left(x^2 + (y-1)^2\right), \quad \text{subject to: } x = y.$$

Projection problem of the point $(0,1)$ on the line $\{(x,y)\,|\,y = x\}$.

Exercise. Verify the following statements.

- Solution: $x^* = (0.5, 0.5)$.
- Penalty function: $Q_c(x,y) = \frac{1}{2}\left(x^2 + (y-1)^2\right) + \frac{c}{2}(y-x)^2$.
- Solution of P_c: $\begin{pmatrix} x_c \\ y_c \end{pmatrix} = \dfrac{1}{1+2c}\begin{pmatrix} c \\ 1+c \end{pmatrix}$.
- There exists a constant M such that for all $c \geq 0$,

$$\|(x_c, y_c) - (\bar{x}, \bar{y})\| \leq M/c.$$

Quadratic penalization

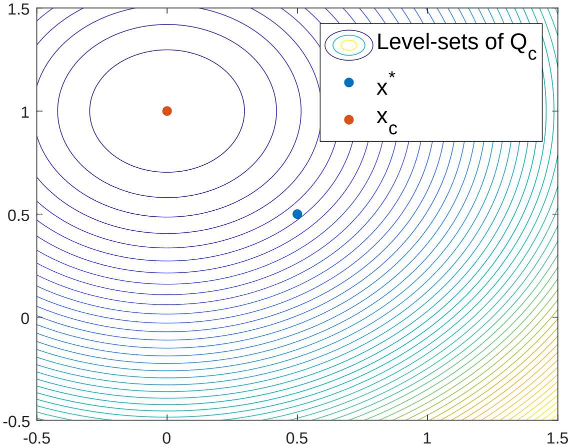


Figure: Graph of Q_c, for $c = 0$.

Quadratic penalization

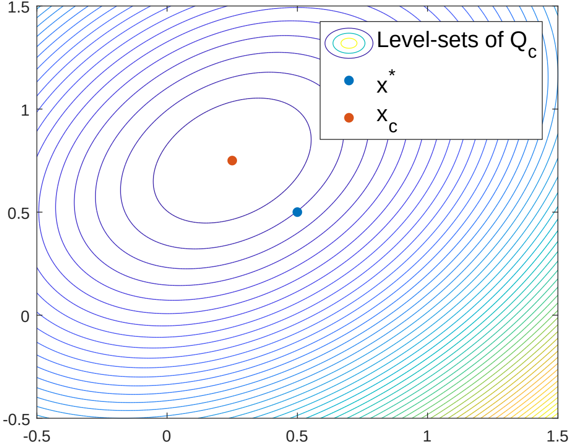


Figure: Graph of Q_c, for $c = 0.5$.

Quadratic penalization

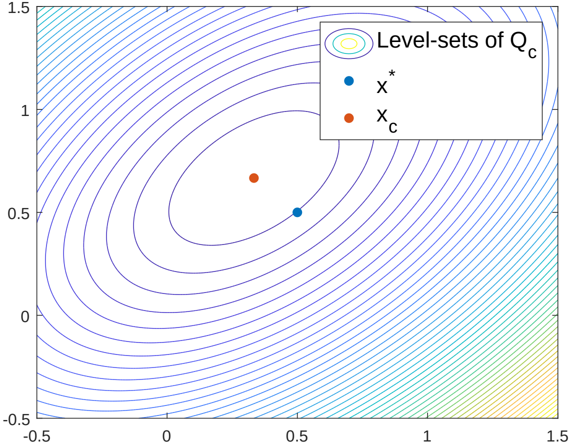


Figure: Graph of Q_c, for $c = 1$.

Quadratic penalization

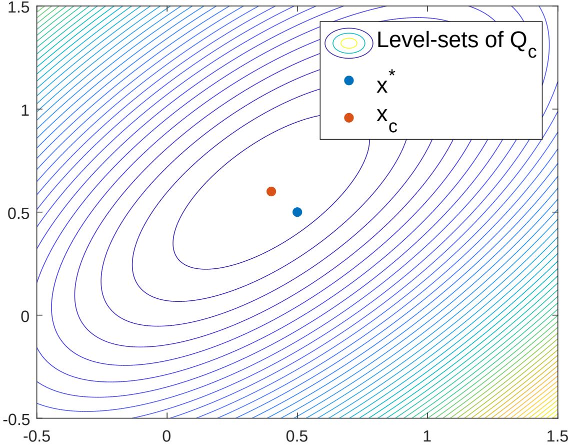


Figure: Graph of Q_c, for $c = 2$.

Quadratic penalization

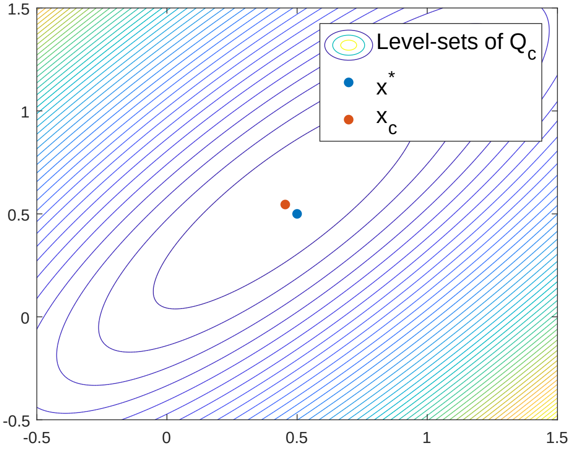


Figure: Graph of Q_c, for $c = 5$.

Quadratic

General idea: increase the value of c progressively, to mitigate the difficulty of minimizing of Q_c.

Algorithm:

1. Input:
   - Number of iterations K
   - Tolerances $\varepsilon_0 > \varepsilon_1 ... > \varepsilon_{K-1} > 0$
   - Penalty parameters $c_0 < c_1 ... < c_{K-1}$
   - Starting point $x_0 \in \mathbb{R}^n$.
2. For $k = 0, ..., K-1$, do
   - Using x_k as a starting point, find x_{k+1} such that

$$\|\nabla Q_{c_k}(x_{k+1})\| \leq \varepsilon_k.$$

   End for.
3. Output: x_K.

Augmented Lagrangian

The two ideas of the **augmented Lagrangian method**:

1. Solving a penalty problem (like (P_c)) also yields an approximation of the Lagrange multiplier.
2. We can "improve" the penalty function Q_c with the knowledge of that approximation.

Algorithm: at each iteration,

- the penalty parameter is increased
- the approximations x_k of the solution and λ_k of the Lagrange multiplier are improved.

Augmented Lagrangian

Idea 1. Let $\phi\colon y \in \mathbb{R}^m \mapsto \frac{1}{2}\|y\|^2 \in \mathbb{R}$. We have $D\phi(y) = y^\top$.

We have $Q_c(x) = f + c\phi \circ g(x)$. By the chain rule we have

$$\begin{aligned} DQ_c(x) &= Df(x) + cD\phi(g(x))Dg(x) \\ &= Df(x) + cg(x)^\top Dg(x) \\ &= D_x L(x, -cg(x)), \end{aligned}$$

where L denotes the Lagrangian.

If $DQ_c(x) \approx 0$, an approximation of the Lagrange multiplier $\bar{\lambda}$ is

$$\bar{\lambda} \approx -cg(x).$$

Augmented Lagrangian

Idea 2. Let $c > 0$. The **augmented Lagrangian** $L_c \colon \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}$ is defined by

$$L_c(x, \lambda) = f(x) - \langle \lambda, g(x) \rangle + \frac{c}{2}\|g(x)\|^2.$$

We have

$$\begin{aligned} L_c(x, \lambda) &= L(x, \lambda) + \frac{c}{2}\|g(x)\|^2 \\ &= Q_c(x) - \langle \lambda, g(x) \rangle \\ &= f(x) + \frac{c}{2}\left\| g(x) - \frac{\lambda}{c} \right\|^2 - \frac{\|\lambda\|^2}{2c}. \end{aligned}$$

Augmented Lagrangian

For a fixed λ, $L_c(\cdot, \lambda)$ still serves as a **penalty function**. If $x_{c,\lambda}$ minimizes $L_c(x, \lambda)$ and if c is very large, then

- $f(x_{c,\lambda})$ is small
- $\frac{c}{2}\|g(x) - \frac{\lambda}{c}\|^2$ is small $\rightarrow \|g(x) - \frac{\lambda}{c}\|$ is very small
  $\rightarrow \|g(x)\|$ is very small.

We also have:

$$\begin{aligned} D_x L_c(x, \lambda) &= Df(x) - \lambda^\top Dg(x) + cg(x)^\top Dg(x) \\ &= D_x L(x, \lambda - cg(x)). \end{aligned}$$

Therefore, if $D_x L_c(x, \lambda) \approx 0$, an approximation of the Lagrange multiplier $\bar{\lambda}$ is

$$\bar{\lambda} \approx \lambda - cg(x).$$

Augmented Lagrangian

The new **penalty problem:**

$$\inf_{x \in \mathbb{R}^n} L_c(x, \lambda). \qquad (P_{c,\lambda})$$

Lemma 2

Let $\bar{x}$ be a local minimizer of (P). Under technical assumptions, there exists $\bar{\lambda}$ and $\bar{c} \geq 0$ such that for all $c > \bar{c}$,

- *the* KKT conditions *hold true*
- *$\bar{x}$ is a* local solution *to $(P_{c,\bar{\lambda}})$.*

Idea of proof. We have

$$D_x L_c(\bar{x}, \bar{\lambda}) = D_x L(\bar{x}, \bar{\lambda} - cg(\bar{x})) = D_x L(\bar{x}, \bar{\lambda}) = 0.$$

For c large enough, $D^2_{xx} L_c(\bar{x}, \bar{\lambda})$ is positive definite. Therefore, $\bar{x}$ is a local solution.

Augmented Lagrangian

Example 1. Consider $\inf_{x\in\mathbb{R}} x - x^2$, subject to: $x = 0$.

- Solution $\bar{x} = 0$.
- Lagrangian $L(x,\lambda) = x - x^2 - \lambda x$. We have

$$D_x L(\bar{x},\lambda) = 1 - 2\bar{x} - \lambda = 1 - \lambda \quad \Longrightarrow \quad \bar{\lambda} = 1.$$

- Augmented lagrangian:

$$L_c(x,\lambda) = x - x^2 - \lambda x + \frac{c}{2}x^2 = (1-\lambda)x + \left(\frac{c}{2} - 1\right)x^2.$$

If $c > \bar{c} := 2$, $L_c(\cdot,\lambda)$ has a unique minimizer

$$x_{c,\lambda} = \frac{\lambda - 1}{c - 2} = \frac{\lambda - \bar{\lambda}}{c - 2}.$$

In particular, $x_{c,\bar{\lambda}} = \bar{x}$.

Augmented Lagrangian

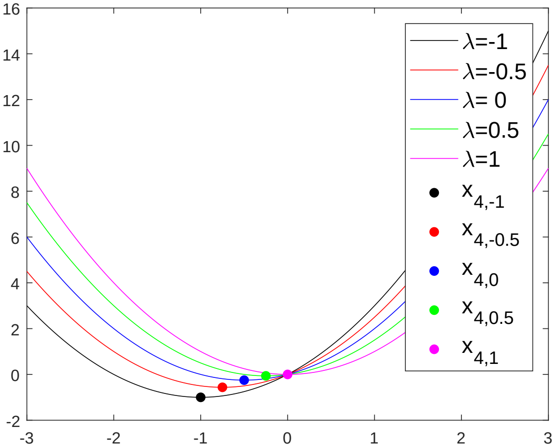


Figure: Graph of $L_c(\cdot, \lambda)$, for $c = 4$ and various values of λ.

Quadratic penalization

Example 2. Consider:

$$\inf_{(x,y)\in\mathbb{R}^2} \frac{1}{2}\big(x^2+(y-1)^2\big), \quad \text{subject to: } x=y.$$

Projection problem of the point $(0,1)$ on the line $\{(x,y)\,|\,y=x\}$.

Exercise. Verify the following statements.

- Solution: $x^* = (0.5, 0.5)$, $\lambda^* = 0.5$. Augmented Lagrangian:

$$L_c(x,y,\lambda) = \tfrac{1}{2}\big(x^2+(y-1)^2\big) - \lambda(x-y) + \tfrac{c}{2}(y-x)^2.$$

- Solution of $(P_{c,\lambda})$: $\begin{pmatrix} x_c \\ y_c \end{pmatrix} = \frac{1}{1+2c}\begin{pmatrix} c+\lambda \\ 1+c-\lambda \end{pmatrix}$.
- There exists a constant M such that for all $c>0$,

$$\|(x_c,y_c)-(\bar{x},\bar{y})\| \leq M|\bar{\lambda}-\lambda|/c.$$

Augmented Lagrangian

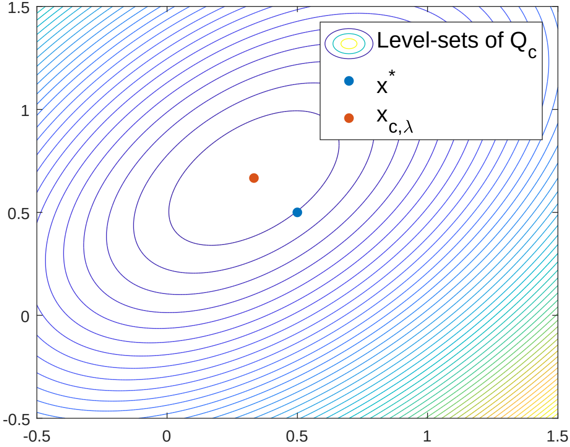


Figure: Level-sets $L_c(\cdot, \lambda)$, for $c = 1$ and $\lambda = 0$.

Augmented Lagrangian

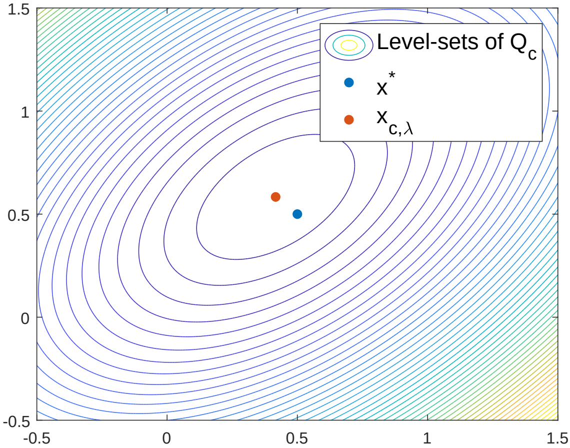


Figure: Level-sets $L_c(\cdot, \lambda)$, for $c = 1$ and $\lambda = 0,25$.

Augmented Lagrangian

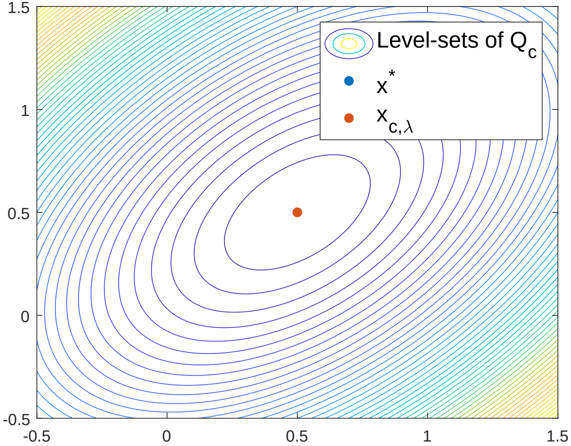


Figure: Level-sets $L_c(\cdot, \lambda)$, for $c = 1$ and $\lambda = 0, 5$.

Augmented Lagrangian

Algorithm.

1. Input:
   - Initial point and multipliers $(x_0, \lambda_0) \in \mathbb{R}^n \times \mathbb{R}^m$
   - Initial penalty parameter $c_0 > 0$, initial tolerance $\varepsilon_0 > 0$
   - Tolerance $\varepsilon > 0$.
2. Set $k = 0$.
3. While $\|D_x L(x_k, \lambda_k)\| > \varepsilon$ and $\|g(x_k)\| > \varepsilon$,
   - Find x_{k+1} such that $\|D_x L_{c_k}(x_{k+1}, \lambda_k)\| \leq \varepsilon_k$.
   - If $\|g(x_{k+1})\|$ is small, set $\lambda_{k+1} = \lambda_k - c_k g(x_{k+1})$. Reduce ε_k.
   - Otherwise, increase c_k.
   - Set $k = k + 1$.

   End while.
4. Output (x_k, λ_k).

Lagrangian decomposition

Main ideas of **Lagrangian decomposition** methods:

- We take $c = 0$ in the augmented Lagrangian. At iterate k, given an approximation λ_k of the Lagrange multiplier, we solve

$$\inf_{x \in \mathbb{R}^n} L(x, \lambda_k). \qquad (P_{\lambda_k})$$

- Given a solution x_{k+1}, the Lagrange multiplier is updated with

$$\lambda_{k+1} = \lambda_k - \alpha g(x_{k+1}),$$

  where $\alpha > 0 \rightarrow$ **Uzawa's algorithm**.

- The update of λ_k is an **ascent gradient step** for a maximization problem related to (P), called dual problem.

Lagrangian decomposition

Remarks.

- Convergence of such methods can be established only under **convexity assumptions**.
- The stepsize $\alpha > 0$ must in general be small enough to ensure convergence. Instead of a fixed stepsize, one can use

$$\lambda_{k+1} = \lambda_k - \alpha_k g(x_{k+1}),$$

where $\alpha_k > 0$ and where

$$\sum_{k=0}^{\infty} \alpha_k = \infty \quad \text{and} \quad \sum_{k=0}^{\infty} \alpha_k^2 = 0.$$

Lagrangian decomposition

Main advantage of Lagrangian decomposition: very often the minimization of L can be “parallelized”.

Standard case: additive constraints.

- Consider

$$\inf_{(x_1,x_2)\in X_1\times X_2} f_1(x_1)+f_2(x_2), \quad \text{subject to: } g_1(x_1)+g_2(x_2)=d,$$

  where f_1, f_2, X_1, X_2, g_1, g_2, and d are given.

- Lagrangian:

$$\begin{aligned} L(x_1,x_2,\lambda) &= f_1(x_1)+f_2(x_2)+\langle\lambda, g_1(x_1)+g_2(x_2)-d\rangle \\ &= \Big[\underbrace{f_1(x_1)+\langle\lambda, g_1(x_1)\rangle}_{=:L_1(x_1,\lambda)}\Big]+\Big[\underbrace{f_2(x_2)+\langle\lambda, g_2x_2\rangle}_{=:L_2(x_2,\lambda)}\Big]-\langle\lambda, d\rangle. \end{aligned}$$

Lagrangian decomposition

Given λ, the minimization of $L(\cdot, \lambda)$ is **decomposed** into two subproblems:

$$\inf_{x_1 \in \mathbb{R}^{n_1}} L_1(x_1, \lambda) \qquad \text{and} \qquad \inf_{x_2 \in \mathbb{R}^{n_2}} L_2(x_2, \lambda),$$

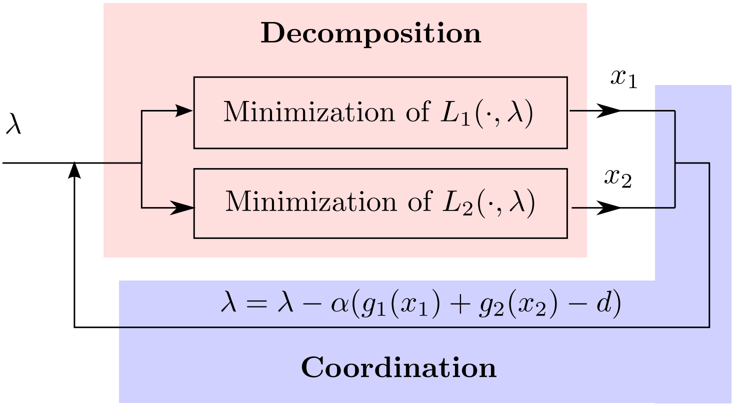
which can be solved independently. Very often the two subproblems are **much easier** to solve than the original problem.

Remark. Straightforward **generalization** to the case

$$\inf_{\substack{x_1, \ldots, x_K \\ \in \mathbb{R}^{n_1} \times \ldots \mathbb{R}^{n_K}}} f_1(x_1) + \ldots + f_K(x_K), \quad \text{s.t.: } g_1(x_1) + \ldots + g_K(x_K) = K.$$

$\rightarrow$ Decomposition in K subproblems (at each iteration).

Lagrangian decomposition



Lagrangian decomposition

1. **Application 1:** space decomposition.
   - **Two production units**, with two independent production processes represented by the variables and constraints
   $$x_1 \in X_1 \quad \text{and} \quad x_2 \in X_2.$$
   - Costs: $f_1(x_1)$ and $f_2(x_2)$.
   - Production of each unit: $g_1(x_1)$ and $g_2(x_2)$.
   - **Coupling constraint:** total production = demand d, that is:
   $$g_1(x_1) + g_2(x_2) = d.$$

Lagrangian decomposition

Interpretation of the decomposition mechanism.

- A "centralizer" buys the production of each unit at a **unitary price** λ_k.
- Each unit must optimize: $\inf_{x_i \in X_i} f_i(x_i) - \lambda_k g_i(x_i)$.
  Solution: $x_{i,k+1}$.
- If $g_1(x_{1,k+1}) + g_2(x_{2,k+1}) < d$, the total prod. is too small.
  $\rightarrow$ The incentive λ to produce is too small.
  It must be increased.
  If $g_1(x_{1,k+1}) + g_2(x_{2,k+1}) > d$, the total prod. is too large.
  $\rightarrow$ The incentive λ to produce is too big.
  It must be decreased.
- This is **consistent** with the formula

$$\lambda_{k+1} = \lambda_k - \alpha_{k+1}\big(g_1(x_{1,k+1}) + g_2(x_{2,k+1}) - d\big).$$

Lagrangian decomposition

Application 2: time decomposition.

- A production process is decomposed over **two periods**.
- Optimization variables:
  - x_1: decisions specific to period 1
  - x_2: decisions specific to period 2
  - y: decisions involved at both periods.

Example: management of a **stock of gas**.

- Quantity withdrawn from the stock at period i: x_i.
- Level of stock between period 1 and period 2: y.

Lagrangian decomposition

- Abstract problem:

$$\inf_{\substack{(x_1,x_2,y)\\(x_1,y)\in X_1\\(x_2,y)\in X_2}} f_1(x_1,y)+f_2(x_2,y).$$

- Equivalent "**decomposable**" problem:

$$\inf_{\substack{(x_1,x_2,y_1,y_2)\\(x_1,y_1)\in X_1\\(x_2,y_2)\in X_2}} f_1(x_1,y_1)+f_2(x_2,y_2), \quad \text{subject to: } y_2-y_1=0.$$

- Independent (w.r.t. time) sub-problems:

$$\inf_{(x_1,y_1)\in X_1} f_1(x_1,y_1)+\lambda_k y_1 \quad \text{and} \quad \inf_{(x_2,y_2)\in X_2} f_2(x_2,y_2)-\lambda_k y_2.$$

Lagrangian decomposition

Application 3: stochastic decomposition.

- A production process is decomposed over two periods. A **random event** with two outcomes ω_1 and ω_2, with probabilities p and $(1-p)$, arises inbetween.
- Optimization variables:
  - x_1: decisions taken if outcome ω_1 arises
  - x_2: decisions taken if outcome ω_2 arises
  - y: decisions taken before the random event.

  Example: purchase of gas y on a day-ahead market (that is, on a given day for the next one).
  Random event: temperature, which impacts consumption.

Lagrangian decomposition

- Abstract problem:

$$\inf_{\substack{(x_1,x_2,y)\\(x_1,y)\in X\\(x_2,y)\in X}} pf(x_1,y,\omega_1)+(1-p)f(x_2,y,\omega_2).$$

- Equivalent problem (with non-anticipativity constraint):

$$\inf_{\substack{(x_1,x_2,y_1,y_2)\\(x_1,y_1)\in X\\(x_2,y_2)\in X}} pf(x_1,y_1,\omega_1)+(1-p)f(x_2,y_2,\omega_2), \quad \text{s.t. } y_2-y_1=0.$$

- Independent (w.r.t. randomness) sub-problems:

$$\inf_{(x_1,y_1)\in X_1} pf_1(x_1,y_1,\omega_1)+\lambda_k y_1, \qquad \inf_{(x_2,y_2)\in X_2} (1-p)f_2(x_2,y_2,\omega_2)-\lambda_k y_2.$$

1. Penalty methods for constrained optimization
   - Quadratic penalization
   - Augmented Lagrangian
   - Lagrangian decomposition

2. Projected gradient method
   - Projection
   - Method
   - Combination with penalty methods

3. Interior point method
   - Nonnegative variables
   - General problems

Projection

The projected gradient method uses a mapping called **projection** defined below.

Lemma 3

Let $K \subset \mathbb{R}^n$ be a non-empty, convex, and closed set. For all $x_0 \in \mathbb{R}^n$, there exists a unique solution *to the problem*

$$\inf_{x \in \mathbb{R}^n} \|x - x_0\|^2, \quad \textit{subject to: } x \in K.$$

It is called projection *of x_0 on K, and denoted $Proj_K(x_0)$.*

Remark. The projection depends on the chosen norm $\|\cdot\|$. For simplicity, we consider the Euclidean norm.

Projection

Example 1: projection on a cuboid.
Let K be described by

$$K = \{x \in \mathbb{R}^n \,|\, \ell_i \leq x_i \leq u_i\},$$

where the coefficients $\ell_1,...,\ell_n \in \mathbb{R} \cup \{-\infty\}$ and $u_1,...,u_n \in \mathbb{R} \cup \{+\infty\}$ are given.

Let $x \in \mathbb{R}^n$, let $y = \text{Proj}_K(x)$. Then

$$y_i = \min(\max(x_i, \ell_i), u_i), \quad \forall i = 1, ..., n.$$

Projection

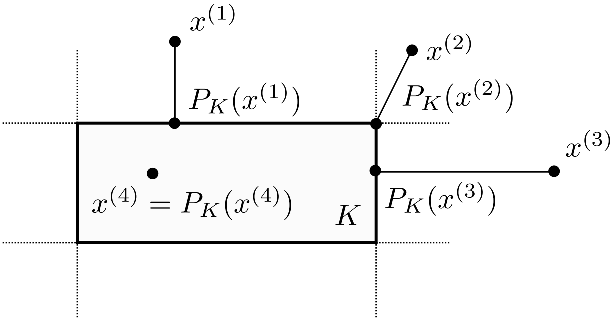


Figure: Projection on a cuboid.

Projection

Example 2: projection on a ball.
Let K be described by

$$K = \left\{x \in \mathbb{R}^n \,|\, \|x - x_C\| \leq R\right\},$$

where $x_C \in \mathbb{R}^n$ and $R \geq 0$ are given.

For all $x \in \mathbb{R}^n$,

$$\mathrm{Proj}_K(x) = x_C + \min\left(\|x - x_C\|, R\right) \frac{(x - x_C)}{\|x - x_C\|}.$$

Projection

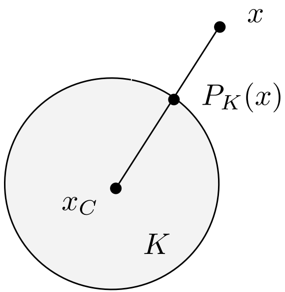


Figure: Projection on a ball.

Projection

Example 3: cartesian product.
Let K be given by

$$K = K_1 \times K_2,$$

where K_1 and K_2 are given non-empty closed and convex subsets of $\mathbb{R}^{n_1}$ and $\mathbb{R}^{n_2}$.

Then for all $x = (x_1, x_2) \in \mathbb{R}^{n_1+n_2}$,

$$\mathrm{Proj}_K(x) = \Big(\mathrm{Proj}_{K_1}(x_1), \mathrm{Proj}_{K_2}(x_2)\Big).$$

Method

Optimization problem. Consider

$$\inf_{x \in \mathbb{R}^n} f(x), \quad x \in K,$$

where $f : \mathbb{R}^n \to \mathbb{R}$ is given and differentiable and K is a given non-empty **convex** and **closed** subset of $\mathbb{R}^n$.

Numerical assumption: $\mathrm{Proj}_K(\cdot)$ is **easy to compute**.

Main idea: at iteration k, replace the search on the half line $\{x_k - \alpha \nabla f(x_k) \,|\, \alpha \geq 0\}$ used in unconstrained optimization by a **search** on

$$\Big\{ \underbrace{\mathrm{Proj}_K\big(x_k - \alpha \nabla f(x_k)\big)}_{=: x_k(\alpha)} \,\Big|\, \alpha \geq 0 \Big\}.$$

Method

Technical aspects.

- Armijo's rule is replaced by the rule

$$f(x_k(\alpha)) \leq f(x_k) + c_1 \langle \nabla f(x_k), x_k(\alpha) - x_k \rangle.$$

- The backstepping method can be used in this context.
- A possible stopping criterion is

$$\|x_k(1) - x_k\| \leq \varepsilon.$$

Combination with penalty methods

Consider the problem

$$\inf_{x\in\mathbb{R}^n} f(x), \quad \text{subject to: } \begin{cases} g_i(x) = 0 & \forall i \in \mathcal{E}, \\ g_i(x) \geq 0 & \forall i \in \mathcal{I}, \end{cases}$$

where $f:\mathbb{R}^n \to \mathbb{R}$ and $g:\mathbb{R}^n \to \mathbb{R}^m$ are given and $(\mathcal{E},\mathcal{I})$ is a partition of $\{1,\dots,m\}$.

An **equivalent formulation** is

$$\inf_{\substack{x\in\mathbb{R}^n\\ y\in\mathbb{R}^m}} f(x), \quad \text{subject to: } \begin{cases} g(x) - y = 0 \\ y \in K, \end{cases}$$

where: $K = \left\{ y \in \mathbb{R}^m \,\middle|\, \begin{cases} y_i = 0 & \forall i \in \mathcal{E} \\ y_i \geq 0 & \forall i \in \mathcal{I} \end{cases} \right\}.$

Combination with penalty methods

Main idea: projection on K (a cuboid) is easy to compute.
Handle $y \in K$ with the projected gradient method.

Algorithm.

- At iteration k, the iterates $x_k \in \mathbb{R}^n$, $y_k \in \mathbb{R}^m$, $\lambda_k \in \mathbb{R}^m$, and c_k are given.
- Solve (approximately) the penalty problem:

$$\inf_{\substack{x \in \mathbb{R}^n \\ y \in \mathbb{R}^m}} L_{c_k}(x, y, \lambda_k) := f(x) - \langle \lambda_k, g(x) - y \rangle + \frac{c_k}{2} \|g(x) - y\|^2,$$

$$\text{subject to: } y \in K,$$

  with the projected gradient method.
  Use (x_k, y_k) as a starting point.

Nonnegative variables

Optimization problem. Consider

$$\inf_{x \in \mathbb{R}^n} f(x), \quad \text{subject to: } x_i \geq 0, \ \forall i = 1, \ldots, n.$$

Barrier function: given $c > 0$, consider the function B_c defined by

$$B_c(x) = f(x) - c \sum_{i=1}^{n} \ln(x_i),$$

for all $x \in \mathbb{R}^n_{>0} := \{y \in \mathbb{R}^n \,|\, y_i > 0, \ \forall i = 1, \ldots, n\}$.

Main idea: approximate (P) by

$$\inf_{x \in \mathbb{R}^n_{>0}} B_c(x). \qquad (P_c)$$

Nonnegative variables

General comments.

- We have: $-\ln(x_i) \to \infty$ as $x_i \to 0$.
  $\to$ Feasible points close to the boundary of the feasible set are **penalized** (whatever the value of c).
- A strong modification of the cost function on the feasible set is undesirable.
  $\to$ The barrier parameter c should be ideally **very small**.
- Problem (P_c) can be solved with methods for **unconstrained optimization**.
  The standard stepsize rules (Armijo,...) prevents us from getting to close to the boundary.
  Ill-conditioning for small values of c.

Nonnegative variables

Example 1.

Consider

$$\inf_{x \in \mathbb{R}} x, \quad \text{subject to: } x \geq 0.$$

- Solution: $\bar{x} = 0$.
- Barrier function: $B_c(x) = x - c\ln(x)$.

$$\nabla B_c(x) = 1 - \frac{c}{x} = 0 \iff x = c.$$

  Since B_c is convex, $x_c := c$ is the global solution to (P_c).
- We have $x_c \underset{c \to 0}{\longrightarrow} 0$.

Nonnegative variables

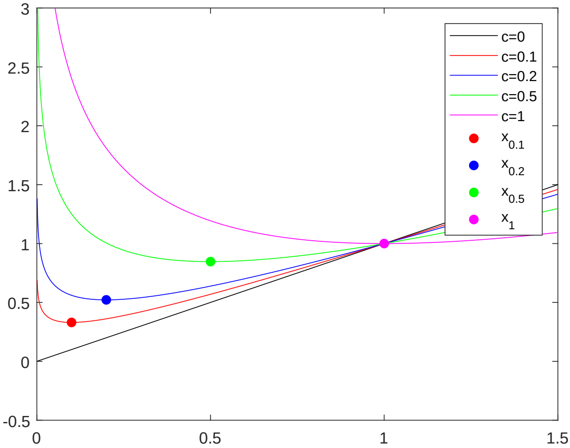


Figure: Level-sets $B_c(\cdot)$, for various values of c

Nonnegative variables

Example 2.
Consider

$$\inf_{(x,y)\in\mathbb{R}^2} \frac{1}{2}(y-1)^2 + x, \quad \text{subject to: } \begin{cases} x \geq 0 \\ y \geq 0. \end{cases}$$

- Solution: $(\bar{x}, \bar{y}) = (0, 1)$.
- Solution to the barrier problem: $(x_c, y_c) = \left(c, \dfrac{1+\sqrt{1+4c}}{2}\right)$.

Nonnegative variables

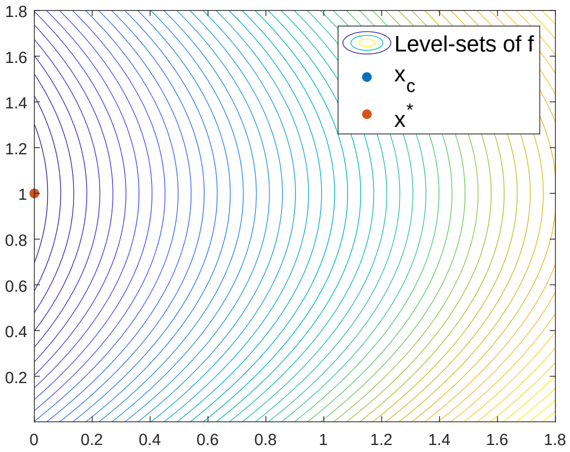


Figure: Level-sets of $f(\cdot)$

Nonnegative variables

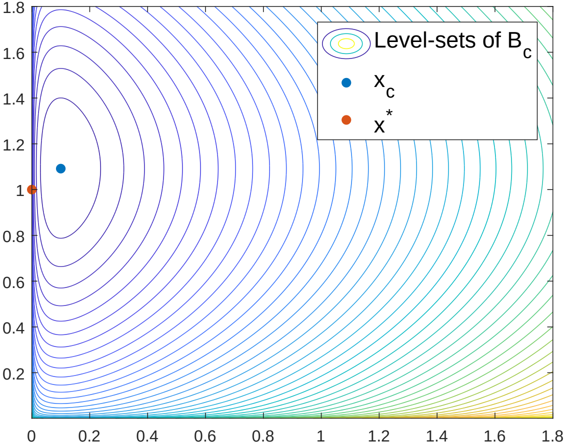


Figure: Level-sets of f, for $c = 0.1$

Nonnegative variables

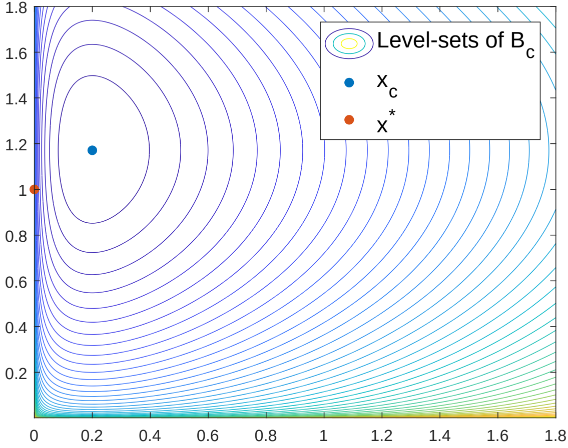


Figure: Level-sets of $B_c(\cdot)$, for $c = 0.2$

Nonnegative variables

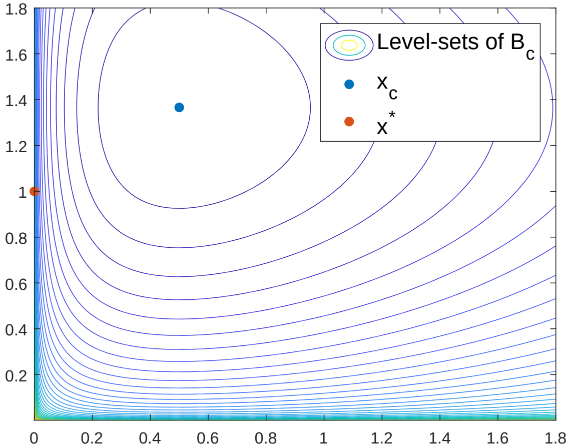


Figure: Level-sets of $B_c(\cdot)$, for $c = 0.5$

Nonnegative variables

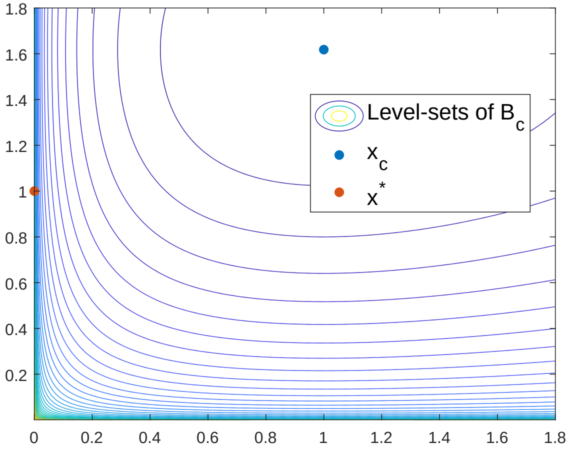


Figure: Level-sets of $B_c(\cdot)$, for $c = 1$

Nonnegative variable

Interpretation with the KKT conditions.

Let $\bar{x}$ be a solution to (P). Let $\bar{\lambda} \in \mathbb{R}^n$ be the associated Lagrange multiplier.

- Lagrangian: $L(x, \lambda) = f(x) - \langle \lambda, x \rangle$. Stationarity condition:

$$\nabla_x L(\bar{x}, \bar{\lambda}) = \nabla f(\bar{x}) - \bar{\lambda} = 0.$$

- Sign condition: $\bar{\lambda}_i \geq 0$.
- Complementarity condition: $\bar{x}_i > 0 \Longrightarrow \bar{\lambda}_i = 0$.
  Equivalently: $\bar{x}_i \bar{\lambda}_i = 0$.

Nonnegative variable

Optimality conditions for the barrier problem.

- For any $x \in \mathbb{R}^n_{>0}$ we denote $\frac{1}{x} = \left(\frac{1}{x_1}, \ldots, \frac{1}{x_n}\right)$.
- Let x_c be a solution to (P_c). We have

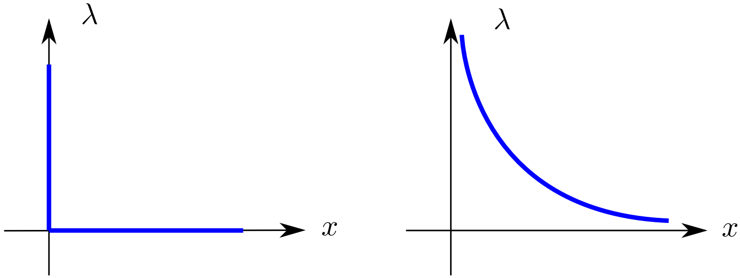
$$\frac{\partial B_c}{\partial x_i}(x_c) = \frac{\partial f}{\partial x_i}(x_c) - \frac{c}{x_i} = 0.$$

  Therefore $\nabla B_c(x_c) = \nabla f(x_c) - \dfrac{c}{x_c} = \nabla_x L(x_c, \dfrac{c}{x_c})$.

- Define $\lambda_c = \frac{c}{x_c} \in \mathbb{R}^n_{>0}$.
  The pair (x_c, λ_c) satisfies the KKT conditions approximately:

$$\nabla L(x_c, \lambda_c) = 0, \quad x_{c,i}\lambda_{c,i} = c, \ \forall i \in \{1, \ldots, n\}.$$

Nonnegative variables



$$\{(x, \lambda) \,|\, x \geq 0,\ \lambda \geq 0,\ x\lambda = 0\} \qquad \{(x, \lambda) \,|\, x \geq 0,\ \lambda \geq 0,\ x\lambda = c\}$$

Figure: Regularization of the complementarity condition

General problem

Consider

$$\inf_{x\in\mathbb{R}^n} f(x), \quad \begin{cases} g_E(x) = 0 \\ g_I(x) \geq 0. \end{cases}$$

Equivalent formulation with slack variable:

$$\inf_{x\in\mathbb{R}^n} f(x), \quad \begin{cases} g_E(x) = 0 \\ g_I(x) - y = 0 \\ y \geq 0. \end{cases}$$

General problems

Numerical approaches.

Approach 1:

- Replace $y \geq 0$ by a barrier term: $-c \sum_{i=1}^{n} \ln(y_i)$.
- **Penalize** the remaining equality constraints.

Approach 2:

- Replace $y \geq 0$ by a barrier term: $-c \sum_{i=1}^{n} \ln(y_i)$.
- Apply **Newton's method** to the KKT conditions associated with the obtained problem.