Continuous optimization
ENT 305

Elise Grosjean

Ensta-Paris
Institut Polytechnique de Paris
2025

And to sum up the courses ...

	Necessary conditions	Sufficient conditions
Abstract formulation (exist.)		if K compact, $f \in C^0(K)$ then at least one solution
		if K closed, $f \in C^0(K)$, coercive then at least one solution

	Necessary conditions	Sufficient conditions
No constraints $K = \mathbb{R}^d$ (opt.)	if $\overline{x}$ local sol., $f \in C^2(K)$ then, $D^2 f(\overline{x})$ is positive semi-def.	if $f \in C^2(K)$, $\nabla f(\overline{x}) = 0$, $D^2 f(\overline{x})$ positive def. then $\overline{x}$ local sol.
Affine constraints	$\overline{x}$ local sol. then KKT	f convex, then KKT=global sol.
Non-linear constraints	$\overline{x}$ local sol., LICQ then KKT	f convex, h affine, g convex, then KKT=global sol.

And to sum up the courses ...

	Necessary conditions	Sufficient conditions
Abstract formulation (exist.)		if K compact, $f \in C^0(K)$ then at least one solution
		if K closed, $f \in C^0(K)$, coercive then at least one solution

	Find a local solution
No constraints	Gradient Descent
Affine constraints	Penalty methods
Non-linear constraints	

Introduction

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

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

well suited if constraints projection is possible and easy to compute.

Reference:

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

Boyd and Vandenberghe. Convex Optimization. Cambridge University Press, 2004.

Quadratic penalization

We consider in this section

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

where $f : \mathbb{R}^n \to \mathbb{R}$ and $h : \mathbb{R}^n \to \mathbb{R}^m$ 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

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

Quadratic penalization

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

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

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

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

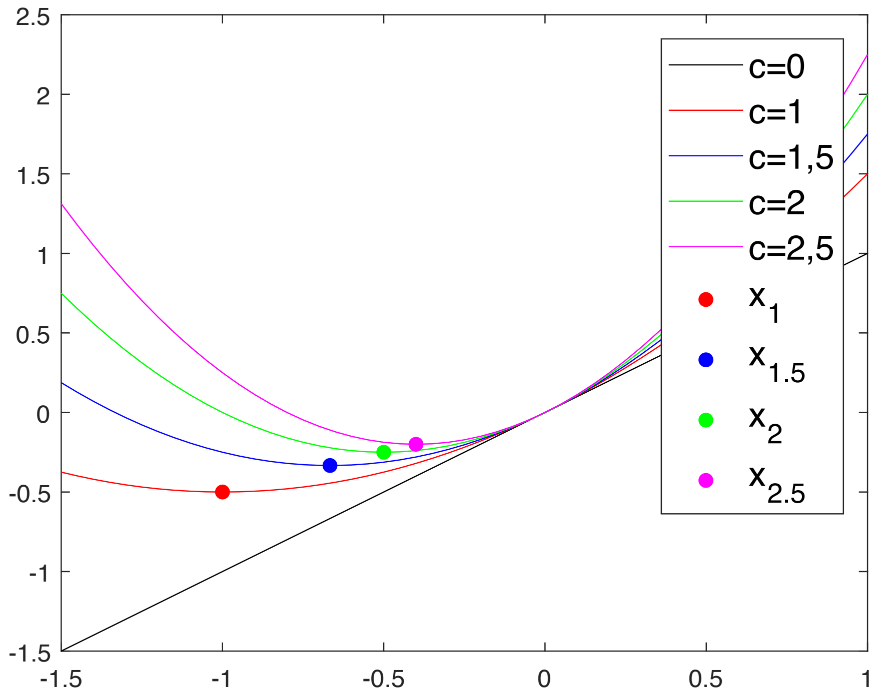


Figure: Graph of Q_c, for various values of c

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 $h(\tilde{x}) = 0$.*

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

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

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

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

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

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 $c_k \|h(x_k)\|^2 \leq M$. Thus

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

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

Quadratic penalization

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

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

Passing to the limit, we get

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

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.
- 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: $x^* = (0.5, 0.5)$.
- Solution of P_c, the penalty function, is:
  $\begin{pmatrix} x_c \\ y_c \end{pmatrix} = \frac{1}{1+2c}\begin{pmatrix} c \\ 1+c \end{pmatrix}.$
- 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

Solution.

1. $\nabla f(x, y) = \begin{pmatrix} x \\ y-1 \end{pmatrix}$. The function f is convex and thus, the global solution of the unconstrained version is $(0, 1)$. With the constraints, we aim at minimizing $\frac{1}{2}(2x^2 - 2x + 1)$, and the unique solution is obviously $x = 0.5$.
2. $Q_c(x) = \frac{1}{2}(x^2 + (y-1)^2) + \frac{c}{2}(y-x)^2$ and $\nabla Q_c(x, y) = \begin{pmatrix} x - c(y-x) \\ y - 1 + c(y-x) \end{pmatrix}$, and since Q_c is convex, the unique solution of P_c is: $\begin{pmatrix} x_c \\ y_c \end{pmatrix} = \dfrac{1}{1+2c} \begin{pmatrix} c \\ 1+c \end{pmatrix}$.
3. $\lim\limits_{c\to\infty} \begin{pmatrix} x_c \\ y_c \end{pmatrix} = \lim\limits_{c\to\infty} \frac{c}{c(1/c+2)} \begin{pmatrix} 1 \\ 1/c+1 \end{pmatrix} = \frac{1}{2}\begin{pmatrix} 1 \\ 1 \end{pmatrix}$
   $\|(x_c, y_c) - (0.5, 0.5)\|^2 = \frac{0.5}{(1+2c)^2} \Rightarrow \|(x_c, y_c) - (0.5, 0.5)\| = \frac{\sqrt{0.5}}{1+2c} \le \frac{M}{c}$.
   Yet, $\nabla^2 Q(x, y) = \begin{pmatrix} 1+c & -c \\ -c & 1+c \end{pmatrix}$ which is ill-conditioned for large c. It yields difficulties with e.g. Newton algorithm $(\nabla^2 Q \cdot p = -\nabla Q)$ with abrupt function changes.

Quadratic penalization

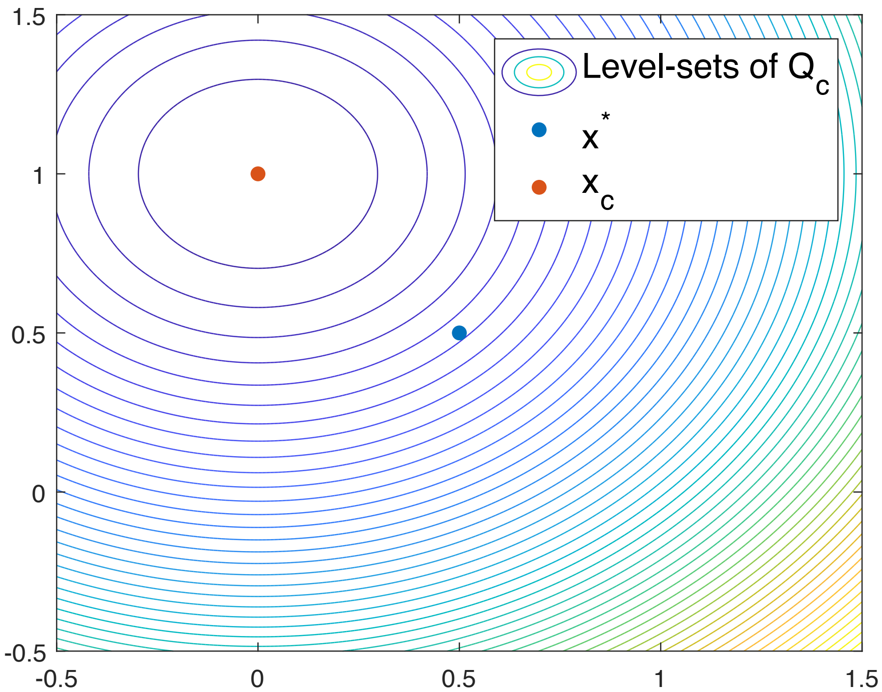


Figure: Graph of Q_c, for $c = 0$.

Quadratic penalization

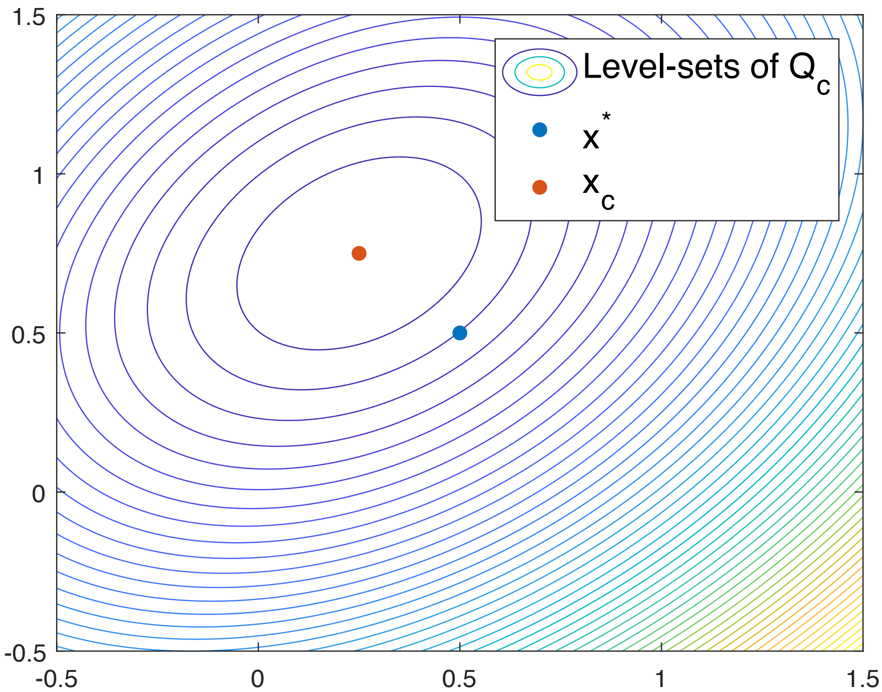


Figure: Graph of Q_c, for $c = 0.5$.

Quadratic penalization

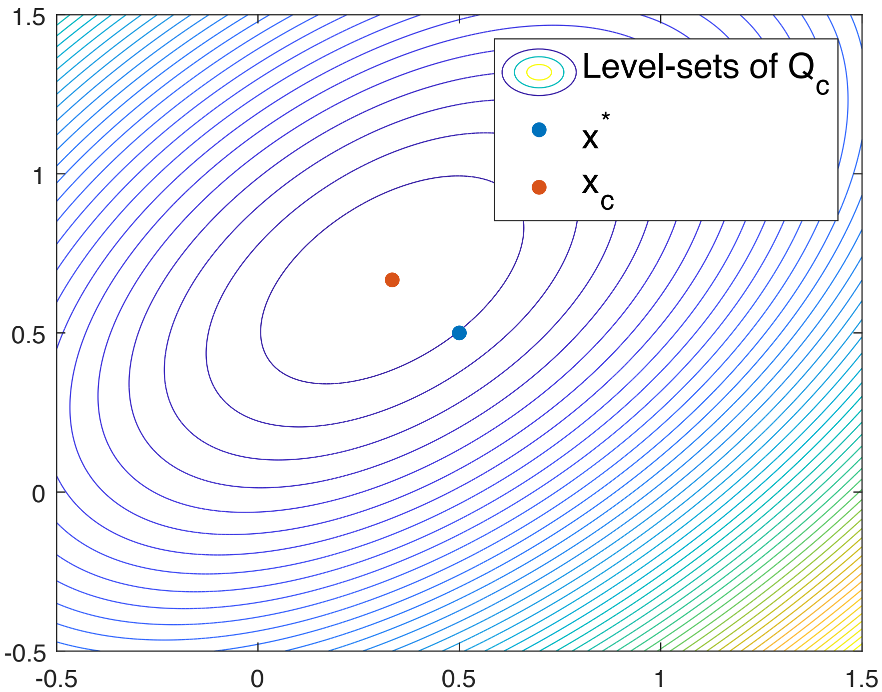


Figure: Graph of Q_c, for $c = 1$.

Quadratic penalization

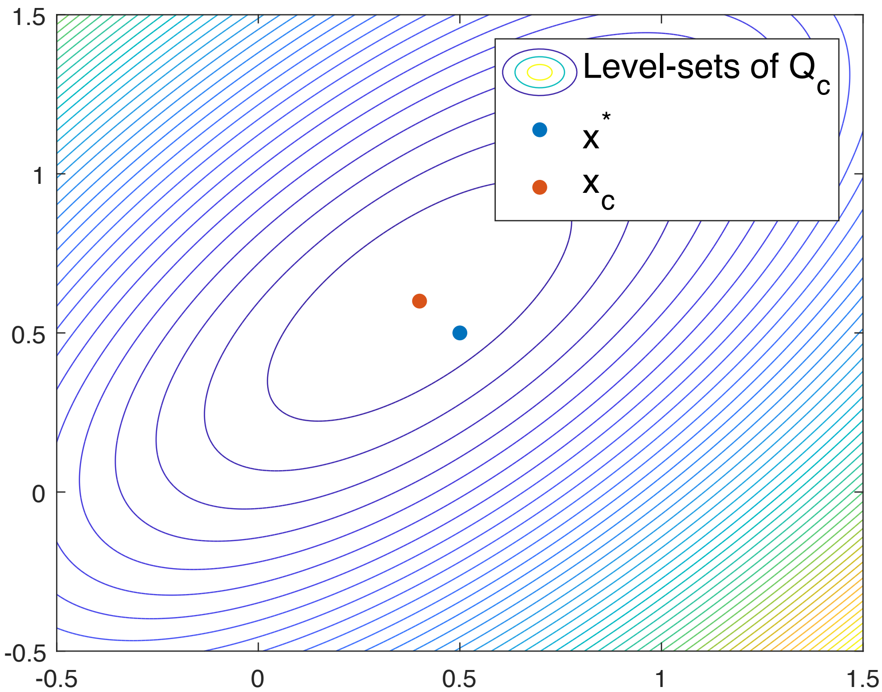


Figure: Graph of Q_c, for $c = 2$.

Quadratic penalization

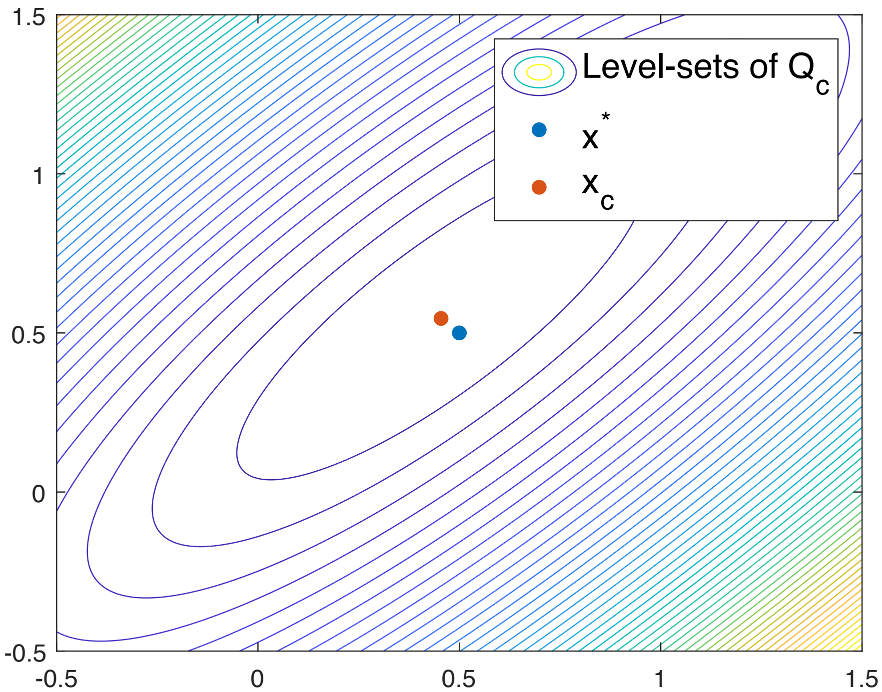


Figure: Graph of Q_c, for $c = 5$.

Penalty algorithm

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

Algorithm:

1. Input: Choose $c_0 > 0$, starting point $x_0 \in \mathbb{R}^n$.
2. For $k = 1, ..., K - 1$, do
   - Solve (P_{c_k}) (e.g. with a gradient descent algorithm starting from x_{k-1}) and set x_k the solution.
   - If x_k is such that $h(x_k) = 0$, stop.
   - Otherwise choose $c_{k+1} > c_k$.

   End for.
3. Output: x_K.

Penalty algorithm

$$Q_c(x) = f(x) + \frac{c}{2}\|h(x)\|^2$$

$$\begin{aligned}\nabla Q_c(x) &= \nabla f(x) + c\langle h(x), \nabla h(x)\rangle \\ &= \nabla L(x, ch(x))\end{aligned}$$

$$c_k h(x_k) \simeq \overline{\mu}$$

Augmented Lagrangian

Unlike the penalty method, with the **augmented Lagrangian method** is not necessary to take $c \to \infty$ in order to solve the original constrained problem, avoiding ill-conditioning.

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

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

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

$$\begin{aligned} \nabla L_c(x, \mu) &= \nabla f(x) + \langle \mu, \nabla h(x) \rangle + \langle c h(x), \nabla h(x) \rangle \\ &= \nabla L(x, \mu + c h(x)) \end{aligned}$$

$$\mu_k + c_k h(x_k) \simeq \overline{\mu}$$

$$h(x_k) \simeq \frac{\overline{\mu} - \mu_k}{c_k}$$

$$\mu_{k+1} = \mu_k + c_k h(x_{k+1})$$

Augmented Lagrangian

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

We have

$$\begin{aligned} L_c(x,\mu) &= L(x,\mu) + \frac{c}{2}\|h(x)\|^2 \\ &= Q_c(x) + \langle \mu, h(x)\rangle \\ &= f(x) + \frac{c}{2}\|h(x) + \frac{\mu}{c}\|^2 - \frac{\|\mu\|^2}{2c} \end{aligned}$$

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

- $f(x_{c,\mu})$ is small
- $\frac{c}{2}\|h(x) + \frac{\mu}{c}\|^2$ is small $\to$ $\|h(x) + \frac{\mu}{c}\|$ is very small
  $\to$ $\|h(x)\|$ is very small.

Augmented Lagrangian

The new **penalty problem:**

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

Lemma 2

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

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

Reminders

	Necessary conditions	Sufficient conditions
Abstract formulation (exist.)		if K compact, $f \in C^0(K)$ then at least one solution
		if K closed, $f \in C^0(K)$, coercive then at least one solution

	Necessary conditions	Sufficient conditions
No constraints $K = \mathbb{R}^d$ (opt.)	if $\overline{x}$ local sol., $f \in C^2(K)$ then, $D^2 f(\overline{x})$ is positive semi-def.	if $f \in C^2(K)$, $\nabla f(\overline{x}) = 0$, $D^2 f(\overline{x})$ positive def. then $\overline{x}$ local sol.
Affine constraints	$\overline{x}$ local sol. then KKT	f convex, then KKT=global sol.
Non-linear constraints	$\overline{x}$ local sol., LICQ then KKT	f convex, h affine, g convex, then KKT=global sol.

Augmented Lagrangian

The new **penalty problem:**

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

Lemma 3

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

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

Idea of proof. We have

$$\nabla L_c(\bar{x}, \bar{\mu}) = \nabla L(\bar{x}, \bar{\mu} + ch(\bar{x})) = \nabla L(\bar{x}, \bar{\mu}) = 0.$$

$$\nabla^2 L_c(\bar{x}, \bar{\mu}) = \nabla^2 L(\bar{x}, \bar{\mu}) + c\langle \nabla h(\bar{x}), \nabla h(\bar{x}) \rangle$$

For c large enough, $\nabla^2 L_c(\bar{x}, \bar{\mu})$ 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: $x = 0$.

Exercise.

- Write the Lagrangian formulation and find the Lagrangian multiplier.
- Does KKT holds for $\bar{x} = 0$?
- Write the augmented Lagrangian $(P_{c,\bar{\mu}})$ and show that $\bar{x}$ is a local solution to $(P_{c,\bar{\mu}})$ if $c > \bar{c}$.

Augmented Lagrangian

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

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

$$\nabla L(\bar{x},\mu) = 1 - 2\bar{x} + \mu = 1 + \mu \quad \Longrightarrow \quad \bar{\mu} = -1.$$

- Augmented lagrangian:

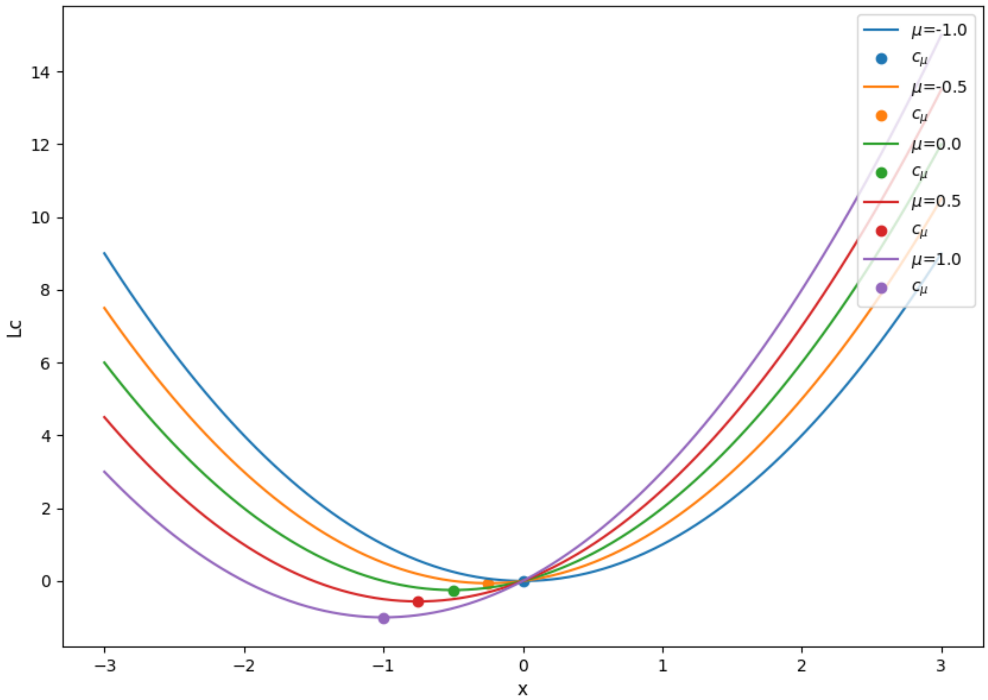
$$L_c(x,\mu) = x - x^2 + \mu x + \frac{c}{2}x^2 = (1+\mu)x + \left(\frac{c}{2} - 1\right)x^2.$$

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

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

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

Augmented Lagrangian



Quadratic penalization

Example 2. 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: $(\bar{x}, \bar{y}) = (0.5, 0.5)$, $\bar{\mu} = 0.5$.
- Solution of $(P_{c,\mu})$ (aug. lagrangian):
  $\begin{pmatrix} x_c \\ y_c \end{pmatrix} = \frac{1}{1+2c} \begin{pmatrix} c+\mu \\ 1+c-\mu \end{pmatrix}$.
- There exists a constant M such that for all $c > 0$,

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

Quadratic penalization

Solution.

1.
- The function f is convex: the global solution of the unconstrained version is given by stationarity:
  $\nabla f(x,y) = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$ which can be rewritten $\begin{pmatrix} x \\ y-1 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$
  We find: $x = 0$ and $y - 1 = 0$ and thus, the global solution of the unconstrained version is $(0,1)$.
- With the constraints $(y = x)$, we can replace y by x in the objective function f: we aim at minimizing $f(x) = \frac{1}{2}\left(2x^2 - 2x + 1\right)$.
  Again, f is convex so the global solution is given by the point satisfying stationarity: $\nabla f(x) = 2x - 1 = 0$. We find the unique solution $\bar{x} = 0.5$.
- To find the Lagrange multiplyer, we replace in the Lagrangian gradient, $\bar{x}$ and $\bar{y}$ by 0.5:
  $L(\bar{x},\bar{y},\bar{\mu}) = f(\bar{x}) + \bar{\mu}(\bar{y} - \bar{x})$, so
  $\nabla L(\bar{x},\bar{y},\bar{\mu}) = \nabla f(\bar{x}) + \bar{\mu}\nabla h(\bar{x},\bar{y}) = \begin{pmatrix} \bar{x} - \bar{\mu} \\ \bar{y} - 1 + \bar{\mu} \end{pmatrix}$ and by stationarity $\nabla L(\bar{x},\bar{y},\bar{\mu}) = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$ implies that $\bar{x} - \bar{\mu} = 0$ and $\bar{y} - 1 + \bar{\mu} = 0$. We find $\bar{\mu} = 0.5$.

Quadratic penalization

2.

$$
\begin{aligned}
L_{c,\mu}(x,y) &= f(x,y) + \frac{c}{2}h(x,y)^2 + \mu h(x,y) \\
&= \frac{1}{2}(x^2 + (y-1)^2) + \frac{c}{2}(y-x)^2 + \mu(y-x)
\end{aligned}
$$

(can be rewritten $\frac{1}{2}(x^2 + (y-1)^2) + \frac{c}{2}(x-y)^2 + \mu(x-y)$) and $\nabla L_{c,\mu}(x,y) = \begin{pmatrix} x - c(y-x) - \mu \\ y - 1 + c(y-x) + \mu \end{pmatrix}$, and since $L_{c,\mu}$ is convex, the unique solution of $(P_{c,\mu})$ is the solution of stationarity condition:
$\nabla L_{c,\mu}(x,y) = \begin{pmatrix} x - c(y-x) - \mu \\ y - 1 + c(y-x) + \mu \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$
It gives two equations: $x - c(y-x) - \mu = 0$ and $y - 1 + c(y-x) + \mu = 0$ Adding the two together, we find $x + y - 1 = 0$, and thus $x = 1 - y$ or $y = 1 - x$.
In the first equation, we replace y by $1 - x$:
$x - c(1 - x - x) - \mu = x(1 + 2c) - c - \mu = 0$, and thus, $x = \frac{c+\mu}{1+2c}$ and in the second one, we replace x by $1 - y$:
$y - 1 + c(y - 1 + y) + \mu = y(1 + 2c) - c - 1 + \mu = 0$, and thus, $y = \frac{1+c-\mu}{1+2c}$. So, $\begin{pmatrix} x_c \\ y_c \end{pmatrix} = \dfrac{1}{1+2c}\begin{pmatrix} c + \mu \\ 1 + c - \mu \end{pmatrix}$.

Quadratic penalization

3. $\|(a,b)\|^2 = a^2 + b^2$ (for euclidean norm)

$$
\begin{aligned}
\|(x_c, y_c) - (\bar{x}, \bar{y})\|^2 &= (\frac{c+\mu}{1+2c} - \bar{x})^2 + (\frac{1+c-\mu}{1+2c} - \bar{y})^2 \\
&= (\frac{c+\mu}{1+2c} - 0.5)^2 + (\frac{1+c-\mu}{1+2c} - 0.5)^2 \\
&= \frac{1}{(1+2c)^2}((c+\mu-0.5-c)^2 + (1+c-\mu-0.5-c)^2) \\
&= \frac{1}{(1+2c)^2}((\mu-0.5)^2 + (0.5-\mu)^2) \\
&= \frac{2(\mu-0.5)^2}{(1+2c)^2} \\
&= \frac{2(\mu-\bar{\mu})^2}{(1+2c)^2}
\end{aligned}
$$

$\|(x_c, y_c) - (0.5, 0.5)\| = \frac{\sqrt{2}}{1+2c}|\mu - \bar{\mu}| \leq \frac{M|\mu-\bar{\mu}|}{c}$.

Augmented Lagrangian

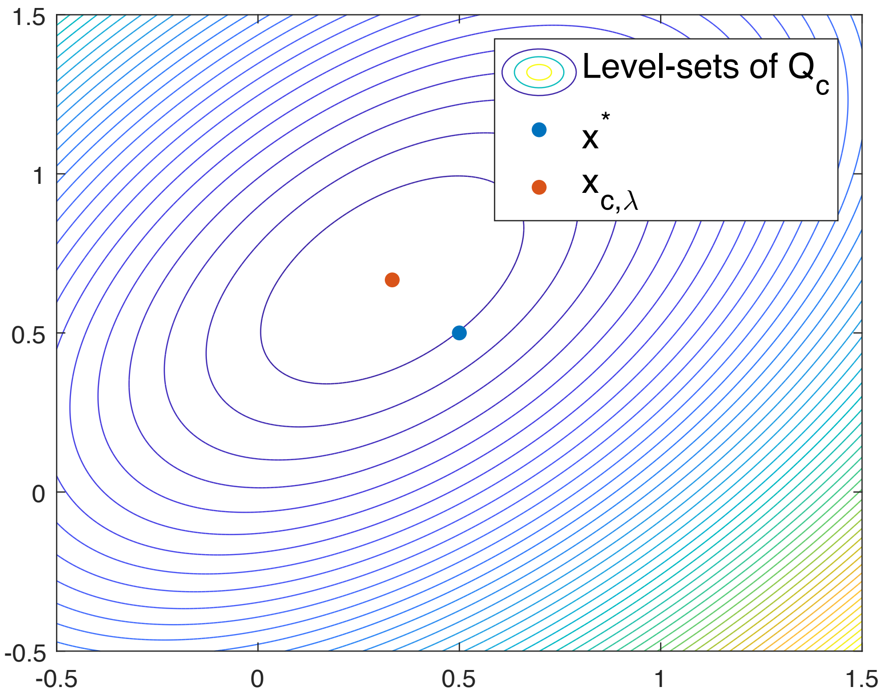


Figure: Level-sets $L_c(\cdot, \mu)$, for $c = 1$ and $\mu = 0$.

Augmented Lagrangian

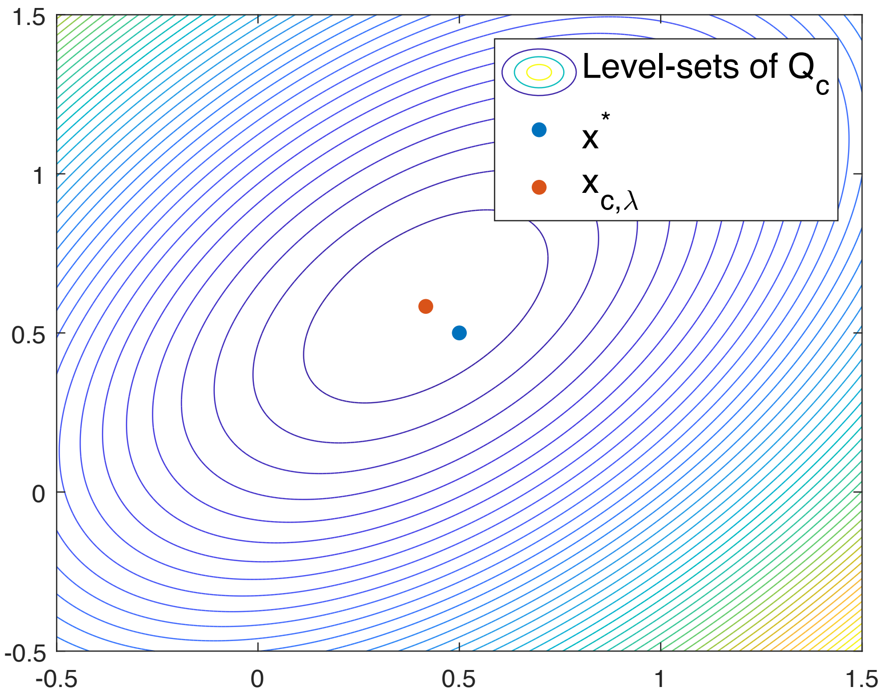


Figure: Level-sets $L_c(\cdot, \mu)$, for $c = 1$ and $\mu = 0, 25$.

Augmented Lagrangian

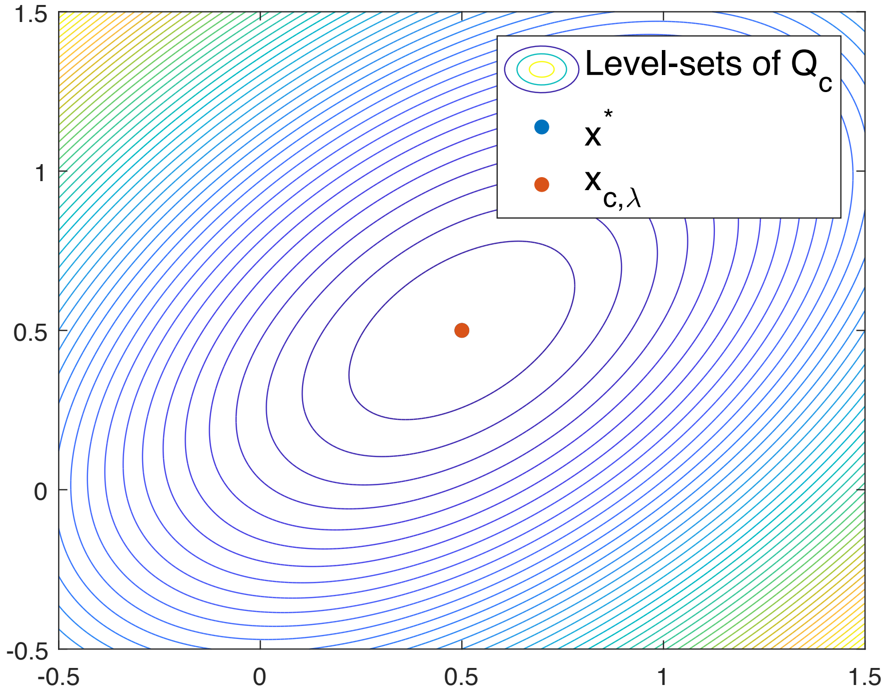


Figure: Level-sets $L_c(\cdot, \mu)$, for $c = 1$ and $\mu = 0, 5$.

Augmented Lagrangian

Algorithm.

1. Input:
   - Initial point and multipliers $(x_0, \mu_0) \in \mathbb{R}^n \times \mathbb{R}^m$
   - Initial penalty parameter $c_0 > 0$, initial tolerance $\varepsilon_0 > 0$
   - Tolerance $\varepsilon > 0$.
2. Set $k = 0$.
3. While $\|D_x L(x_k, \mu_k)\| > \varepsilon$ and $\|h(x_k)\| > \varepsilon$,
   - Find x_{k+1} such that $\|D_x L_{c_k}(x_{k+1}, \mu_k)\| \leq \varepsilon_k$.
   - If $\|h(x_{k+1})\|$ is small, set $\mu_{k+1} = \mu_k + c_k h(x_{k+1})$. Reduce ε_k.
   - Otherwise, increase c_k.
   - Set $k = k + 1$.

   End while.
4. Output (x_k, λ_k).