

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 is non-empty and compact, $f \in C^0(K)$, then at least one sol.
		if K is non-empty and closed, $f \in C^0(K)$, coercive then at least one sol.

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

And to sum up the courses ...

	Necessary conditions	Sufficient conditions
Abstract formulation (exist.)		if K is non-empty and compact, $f \in C^0(K)$, then at least one sol.
		if K is non-empty and closed, $f \in C^0(K)$, coercive then at least one sol.

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

1 Reminders

- KKT
- Sensitivity analysis

2 Penalty methods

- Quadratic penalization
- Augmented Lagrangian

3 Projected gradient method

- Projection
- Method
- Combination with penalty methods

KKT

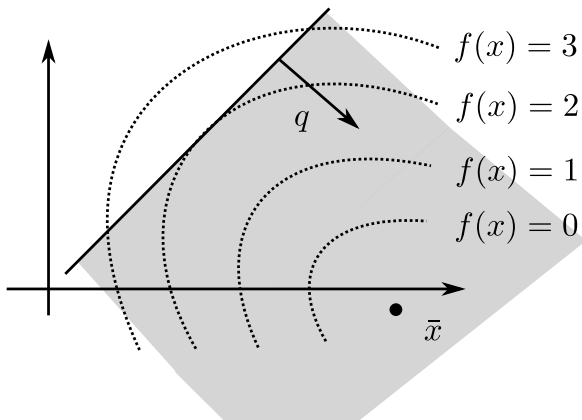
- We investigate in this section the problem

$$\inf_{x \in \mathbb{R}^n} f(x), \quad \text{s.t.} \quad \begin{cases} h_i(x) = 0, & \forall i \in \mathcal{E} \\ g_j(x) \leq 0, & \forall j \in \mathcal{I}. \end{cases} \quad (P)$$

- Let $x \in \mathbb{R}^n$ be feasible. Let $j \in \mathcal{I}$. We say that
 - the inequality constraint j is **active** if $g_j(x) = 0$
 - the inequality constraint j is **inactive** if $g_j(x) < 0$.

Linear constraints

$q = \nabla g$ **Example** Case of one inactive inequality constraint:



- Let $f: \mathbb{R}^n \rightarrow \mathbb{R}$ and let $h: \mathbb{R}^n \rightarrow \mathbb{R}^{m_1}$ and $g: \mathbb{R}^n \rightarrow \mathbb{R}^{m_2}$ be C^2 .
- Let the **Lagrangian** $L: \mathbb{R}^n \times \mathbb{R}^{m_1} \times \mathbb{R}^{m_2} \rightarrow \mathbb{R}$ be defined by

$$\begin{aligned} L(x, \mu, \lambda) &= f(x) + \langle \mu, h(x) \rangle + \langle \lambda, g(x) \rangle \\ &= f(x) + \sum_{i=1}^{m_1} \mu_i h_i(x) + \sum_{j=1}^{m_2} \lambda_j g_j(x). \end{aligned}$$

The variables μ, λ are referred to as **dual variables**.

KKT conditions

Let $\bar{x}$ be feasible for (P) .

The **Karush-Kuhn-Tucker (KKT) conditions** at $\bar{x}$ are the existence of

$$(\mu, \lambda) \in \mathbb{R}^{m_1} \times \mathbb{R}^{m_2}$$

such that:

1 Stationarity:

$$\nabla_x L(\bar{x}, \mu, \lambda) = 0.$$

2 Dual feasibility:

3 Complementarity:

4 Primal feasibility:

KKT conditions

Let $\bar{x}$ be feasible for (P) .

The **Karush-Kuhn-Tucker (KKT) conditions** at $\bar{x}$ are the existence of

$$(\mu, \lambda) \in \mathbb{R}^{m_1} \times \mathbb{R}^{m_2}$$

such that:

1 Stationarity:

$$\nabla_x L(\bar{x}, \mu, \lambda) = 0.$$

2 Dual feasibility:

$$\lambda_j \geq 0, \quad \forall j \in \mathcal{I}.$$

3 Complementarity:

4 Primal feasibility:

KKT conditions

Let $\bar{x}$ be feasible for (P) .

The **Karush-Kuhn-Tucker (KKT) conditions** at $\bar{x}$ are the existence of

$$(\mu, \lambda) \in \mathbb{R}^{m_1} \times \mathbb{R}^{m_2}$$

such that:

1 Stationarity:

$$\nabla_x L(\bar{x}, \mu, \lambda) = 0.$$

2 Dual feasibility:

$$\lambda_j \geq 0, \quad \forall j \in \mathcal{I}.$$

3 Complementarity:

$$\lambda_j g_j(\bar{x}) = 0, \quad \forall j \in \mathcal{I}.$$

4 Primal feasibility:

KKT conditions

Let $\bar{x}$ be feasible for (P) .

The **Karush-Kuhn-Tucker (KKT) conditions** at $\bar{x}$ are the existence of

$$(\mu, \lambda) \in \mathbb{R}^{m_1} \times \mathbb{R}^{m_2}$$

such that:

1 Stationarity:

$$\nabla_x L(\bar{x}, \mu, \lambda) = 0.$$

2 Dual feasibility:

$$\lambda_j \geq 0, \quad \forall j \in \mathcal{I}.$$

3 Complementarity:

$$\lambda_j g_j(\bar{x}) = 0, \quad \forall j \in \mathcal{I}.$$

4 Primal feasibility:

$$h_i(\bar{x}) = 0, \quad \forall i, \quad g_j(\bar{x}) \leq 0, \quad \forall j.$$

Consider the primal problem (P):

$$\inf_{x \in \mathbb{R}^n} f(x) \text{ s.t. } g(x) \leq 0$$

Let p^* be the optimal cost of (P), and L the associated Lagrangian formulation. Then $p^* = \inf_{x \in \mathbb{R}^n} \sup_{\lambda \in \mathbb{R}_+^m} L(x, \lambda)$. Indeed,

$$\begin{aligned} \sup_{\lambda \in \mathbb{R}_+^m} L(x, \lambda) &= \sup_{\lambda \in \mathbb{R}_+^m} \left(f(x) + \sum_{j=1}^m \lambda_j g_j(x) \right), \\ &= \begin{cases} f(x) & \text{if } g_j(x) \leq 0, \forall j = 1, \dots, m \\ \infty & \text{otherwise (if } x \text{ not feasible and } g_j(x) > 0), \end{cases} \end{aligned}$$

Example 4.

Consider the problem

$$\inf_{x \in \mathbb{R}} x, \quad \text{subject to: } x^2 \leq 0.$$

The unique feasible point is $\bar{x} = 0$, hence it is the solution.

Lagrangian:

$$L(x, \lambda) = x + \lambda x^2.$$

At zero:

$$\nabla_x L(0, \lambda) = 1 + 2\lambda\bar{x} = 1 \neq 0.$$

The LICQ is not satisfied, since $\nabla g_1(0) = 0$.

KKT

We consider the problem

$$\min_{x,y} f(x,y) = x$$

subject to

$$g(x, y) = y^2 - x^3 \leq 0.$$

Exercise.

- 1 Write the Lagrangian
- 2 Find $(\bar{x}, \bar{y})$ the global solution
- 3 Are the LICQ conditions satisfied at that point?
- 4 What about KKT conditions?

KKT exercise: solution (nonlinear constraint)

Problem: $\min_{(x,y) \in \mathbb{R}^2} x$ subject to $y^2 - x^3 \leq 0$.

$$L(x, y, \lambda) = x + \lambda(y^2 - x^3), \quad \lambda \geq 0.$$

Feasibility implies $x^3 \geq y^2 \geq 0$, hence $x \geq 0$. The origin is feasible:

(0, 0) is a global minimizer. At (0, 0),

$$\nabla f = (1, 0), \quad \nabla g = (-3x^2, 2y)|_{(0,0)} = (0, 0).$$

Thus **LICQ fails**. Stationarity would require

$$(1, 0) + \lambda(0, 0) = (0, 0),$$

The minimizer does not satisfy KKT.

Sensitivity analysis

- Consider the family of optimization problems

$$\inf_{x \in \mathbb{R}^n} f(x), \quad \text{s.t.} \quad \begin{cases} h_i(x) = y_i, & \forall i \in \mathcal{E}, \\ g_j(x) \leq y_j, & \forall j \in \mathcal{I}, \end{cases} \quad (P(y))$$

parametrized by the vector $y \in \mathbb{R}^m$.

- Let the **value function** V be defined by

$$V(y) = \text{val}(P(y)).$$

- A variation δy_i changes the optimal cost by approximately $-\lambda_i \delta y_i$ (or $-\mu_i \delta y_i$ for an equality constraint).

Sensitivity analysis: active vs. inactive constraints

Consider $Ax \leq b$, with multiplier $\lambda \geq 0$.

Inactive constraint

$$Ax < b$$

Changing the constraint has no local impact.

For a sufficiently small perturbation δb :

$$Ax < b + \delta b.$$

By complementarity:

$$\lambda = 0$$

$$V(b + \delta b) \simeq V(b).$$

Active constraint

$$Ax = b$$

Changing the constraint may affect the optimal cost.

If $V \in \mathcal{C}^1$: $V'(b) = -\lambda$

$$V(b + \delta b) \simeq V(b) - \lambda \delta b.$$

- $\delta b > 0$: relaxing the constraint may reduce the cost.
- $\delta b < 0$: tightening the constraint may increase the cost.

Sensitivity analysis

Exercise.

A company rents a machine for d days and uses it to produce two goods, which cannot be produced simultaneously. Let x_1 and x_2 denote the production times allocated to each good. The resulting profits (in k€) are:

$$\frac{x_1}{1 + x_1} \quad \text{and} \quad \frac{x_2}{4 + x_2}.$$

Sensitivity analysis

- 1 Formulate the problem as a minimization problem.
- 2 Justify the existence of a solution.
- 3 Write the KKT conditions. What is the unit of the dual variable?
- 4 Verify that $\bar{x} = (4, 6)$ satisfies the KKT conditions for $d = 10$ days. Is it a global solution to the problem?
- 5 The rental cost of the machine is 70€/day. Should the company rent the machine for a longer period?

Sensitivity analysis

5. Should the company rent the machine longer?

The optimal production cost satisfies

$$V'(d) = -\mu(d).$$

At $d = 10$, we have $\mu(10) = 0.04$, hence

$$V(10 + \delta d) \simeq V(10) - 0.04 \delta d.$$

The rental cost is 0.07 k€/day. The total optimized cost is therefore

$$C(d) = V(d) + 0.07d.$$

Thus,

$$C'(10) = -\mu(10) + 0.07 = -0.04 + 0.07 = 0.03.$$

Conclusion: An extra rental day generates approximately 40€ in additional production profit, but costs 70€.

The company should not extend the rental period.

1 Reminders

- KKT
- Sensitivity analysis

2 Penalty methods

- Quadratic penalization
- Augmented Lagrangian

3 Projected gradient method

- Projection
- Method
- Combination with penalty methods

Motivation: KKT conditions may fail

Example 1: a nonlinear inequality constraint.

$$\min_{(x,y) \in \mathbb{R}^2} f(x,y) = x, \quad g(x,y) = y^2 - x^3 \leq 0.$$

Since $y^2 \leq x^3$, every feasible point satisfies $x \geq 0$. Thus $(0,0)$ is the **unique global minimizer**.

The Lagrangian is $L(x,y,\lambda) = x + \lambda(y^2 - x^3)$. Stationarity at $(0,0)$ would imply

$$\nabla_{(x,y)} L(0,0,\lambda) = \begin{pmatrix} 1 - 3\lambda \cdot 0^2 \\ 2\lambda \cdot 0 \end{pmatrix} = \begin{pmatrix} 1 \\ 0 \end{pmatrix} \neq 0.$$

- No multiplier $\lambda \geq 0$ satisfies the KKT equations.
- LICQ fails: $\nabla g(0,0) = (0,0)$.

KKT are not always necessary without a constraint qualification.

Motivation: an indefinite KKT system

Example 2: projection onto the line $x = y$.

$$\min_{(x,y) \in \mathbb{R}^2} \frac{1}{2} (x^2 + (y-1)^2), \quad x - y = 0.$$

The Lagrangian is $L(x, y, \mu) = \frac{1}{2}(x^2 + (y-1)^2) + \mu(x - y)$.

$$\underbrace{\begin{pmatrix} 1 & 0 & 1 \\ 0 & 1 & -1 \\ 1 & -1 & 0 \end{pmatrix}}_K \begin{pmatrix} x \\ y \\ \mu \end{pmatrix} = \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}.$$

- The eigenvalues of K are $\boxed{-1, 1, 2}$: K is **indefinite**.
- L is not jointly convex in (x, y, μ) and is unbounded below when minimized over all three variables.
- Standard gradient descent *on L in all variables* is inappropriate: the KKT point is a saddle point.

Why does indefiniteness matter?

For general equality-constrained quadratic programs,

$$\min_{x \in \mathbb{R}^n} \frac{1}{2} x^\top H x - q^\top x, \quad Ax = b, \quad H \succ 0.$$

$$\underbrace{\begin{pmatrix} H & A^\top \\ A & 0 \end{pmatrix}}_K \begin{pmatrix} x \\ \mu \end{pmatrix} = \begin{pmatrix} q \\ b \end{pmatrix}.$$

- Size $(n + m) \times (n + m)$: primal and dual unknowns.
- Standard CG and Cholesky (for SPD matrices) are not directly applicable.
- Rank-deficient A makes this KKT matrix singular when $H \succ 0$.

Remark: indefinite systems can be solved (e.g. MINRES or pivoted $LDL^\top$).
Indefiniteness does *not* imply poor conditioning.

Quadratic penalization: the basic idea

Consider the equality-constrained problem

$$(P) \quad \min_{x \in \mathbb{R}^n} f(x), \quad h(x) = 0,$$

where $f : \mathbb{R}^n \rightarrow \mathbb{R}$ and $h : \mathbb{R}^n \rightarrow \mathbb{R}^m$ are smooth.

Two goals: minimize $f(x)$ and achieve feasibility $h(x) = 0$. At a trial iterate x_k , should we decrease $f(x_k)$ or the constraint violation $\|h(x_k)\|$?

Idea: combine both goals into one objective:

$$Q_c(x) = f(x) + \frac{c}{2} \|h(x)\|^2, \quad c > 0.$$

- Standard unconstrained optimization algorithms can minimize Q_c !
- For finite c , feasibility may only be approximated.
- Increase c progressively to enforce the constraint more closely.

Quadratic penalization: the basic idea

Consider the equality-constrained problem

$$(P) \quad \min_{x \in \mathbb{R}^n} f(x), \quad h(x) = 0,$$

where $f : \mathbb{R}^n \rightarrow \mathbb{R}$ and $h : \mathbb{R}^n \rightarrow \mathbb{R}^m$ are smooth.

Two goals: minimize $f(x)$ and achieve feasibility $h(x) = 0$. At a trial iterate x_k , should we decrease $f(x_k)$ or the constraint violation $\|h(x_k)\|$?

Idea: combine both goals into one objective:

$$Q_c(x) = f(x) + \frac{c}{2} \|h(x)\|^2, \quad c > 0.$$

- Standard unconstrained optimization algorithms can minimize Q_c !
- For finite c , feasibility may only be approximated.
- Increase c progressively to enforce the constraint more closely.

Quadratic penalization: the basic idea

Consider the equality-constrained problem

$$(P) \quad \min_{x \in \mathbb{R}^n} f(x), \quad h(x) = 0,$$

where $f : \mathbb{R}^n \rightarrow \mathbb{R}$ and $h : \mathbb{R}^n \rightarrow \mathbb{R}^m$ are smooth.

Two goals: minimize $f(x)$ and achieve feasibility $h(x) = 0$. At a trial iterate x_k , should we decrease $f(x_k)$ or the constraint violation $\|h(x_k)\|$?

Idea: combine both goals into one objective:

$$Q_c(x) = f(x) + \frac{c}{2} \|h(x)\|^2, \quad c > 0.$$

- Standard unconstrained optimization algorithms can minimize Q_c !
- For finite c , feasibility may only be approximated.
- Increase c progressively to enforce the constraint more closely.

Quadratic penalization: the basic idea

Consider the equality-constrained problem

$$(P) \quad \min_{x \in \mathbb{R}^n} f(x), \quad h(x) = 0,$$

where $f : \mathbb{R}^n \rightarrow \mathbb{R}$ and $h : \mathbb{R}^n \rightarrow \mathbb{R}^m$ are smooth.

Two goals: minimize $f(x)$ and achieve feasibility $h(x) = 0$. At a trial iterate x_k , should we decrease $f(x_k)$ or the constraint violation $\|h(x_k)\|$?

Idea: combine both goals into one objective:

$$Q_c(x) = f(x) + \frac{c}{2} \|h(x)\|^2, \quad c > 0.$$

- Standard unconstrained optimization algorithms can minimize Q_c !
- For finite c , feasibility may only be approximated.
- Increase c progressively to enforce the constraint more closely.

What if the constraint is an inequality?

For $g(x) \leq 0$, **feasible points must not be penalized**. define

$$Q_c(x) = f(x) + \frac{c}{2} [\max(0, g(x))]^2.$$

For several constraints:

$$Q_c(x) = f(x) + \frac{c}{2} \|h(x)\|^2 + \frac{c}{2} \sum_{j=1}^{m_2} [\max(0, g_j(x))]^2.$$

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 $x_c \xrightarrow{c \rightarrow +\infty} \bar{x}$.

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,

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

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

- 3 Obviously

$$x_c = -1/c \xrightarrow{c \rightarrow \infty} 0 = \bar{x}.$$

Quadratic penalization

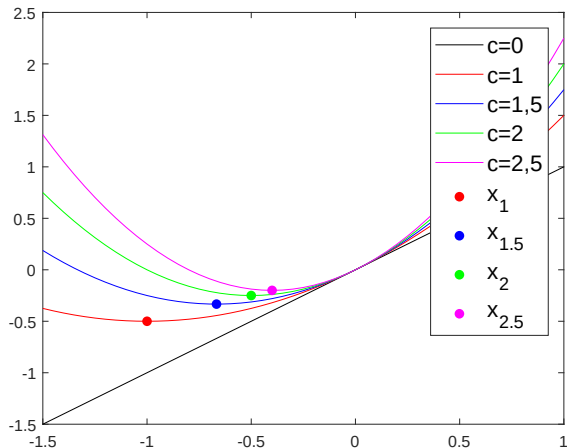


Figure: Graph of Q_c , for various values of c

Quadratic penalty: nonlinear example

Exercise

Consider the constrained optimization problem

$$\min_{(x,y) \in \mathbb{R}^2} f(x,y) = x, \quad y^2 - x^3 \leq 0.$$

For $c > 0$, consider the penalized function

$$Q_c(x,y) = x + \frac{c}{2} [\max(0, y^2 - x^3)]^2.$$

- 1 Recall the global minimizer of the original constrained problem.
- 2 For fixed $x < 0$, show that $Q_c(x,y)$ is minimized with respect to y at $y = 0$.
- 3 Determine a global minimizer (x_c, y_c) of Q_c .
- 4 Show that (x_c, y_c) converges to the constrained minimizer as $c \rightarrow +\infty$.

Quadratic penalty: nonlinear example

Recall $\min_{x,y} x$ subject to $g(x, y) = y^2 - x^3 \leq 0$.

$$Q_c(x, y) = x + \frac{c}{2} [\max(0, y^2 - x^3)]^2.$$

For fixed $x < 0$, $y^2 - x^3 > 0$ and is smallest at $y = 0$. Hence

$$Q_c(x, 0) = x + \frac{c}{2} x^6, \quad \frac{d}{dx} Q_c(x, 0) = 1 + 3cx^5.$$

The unique stationary point on $x < 0$ is

$$x_c = -(3c)^{-1/5}, \quad y_c = 0.$$

- For $x \geq 0$, $Q_c(x, y) \geq x \geq 0$.
- For $x < 0$, the minimum over y occurs at $y = 0$; the value at x_c is negative, so this is the global minimum.
- $(x_c, y_c) \rightarrow (0, 0)$ as $c \rightarrow \infty$.

The penalty method can converge even when KKT fails at the limit.

Quadratic penalization

Lemma 1

Let $c_k \rightarrow \infty$. 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

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

Since $f(x_k) \rightarrow f(\bar{x})$, 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

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

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.
- 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:

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

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

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

Recall $\min_{x,y} \frac{1}{2}(x^2 + (y-1)^2)$ subject to $x = y$. *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 problem 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 stationarity gives

$$\begin{cases} x + c(x-y) = 0, \\ y-1 + c(y-x) = 0. \end{cases}$$

Adding the equations: $x + y = 1$, so $y = 1 - x$. Thus $(1 + 2c)x = c$, and

$$(x_c, y_c) = \left(\frac{c}{1+2c}, \frac{1+c}{1+2c} \right).$$

Quadratic penalization

Recall $\min_{x,y} \frac{1}{2}(x^2 + (y-1)^2)$ subject to $x = y$.

- 1 $\lim_{c \rightarrow \infty} \begin{pmatrix} x_c \\ y_c \end{pmatrix} = \lim_{c \rightarrow \infty} \frac{c}{c(1/c+2)} \begin{pmatrix} 1 \\ 1/c + 1 \end{pmatrix} = \frac{1}{2} \begin{pmatrix} 1 \\ 1 \end{pmatrix}$
 $\|(x_c, y_c) - (0.5, 0.5)\|^2 = \frac{0.5}{(1+2c)^2} \Rightarrow \|(x_c, y_c) - (0.5, 0.5)\| = \frac{\sqrt{0.5}}{1+2c} \leq$
 $\frac{M}{c}$. However, $\nabla^2 Q_c(x, y) = \begin{pmatrix} 1+c & -c \\ -c & 1+c \end{pmatrix}$ 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

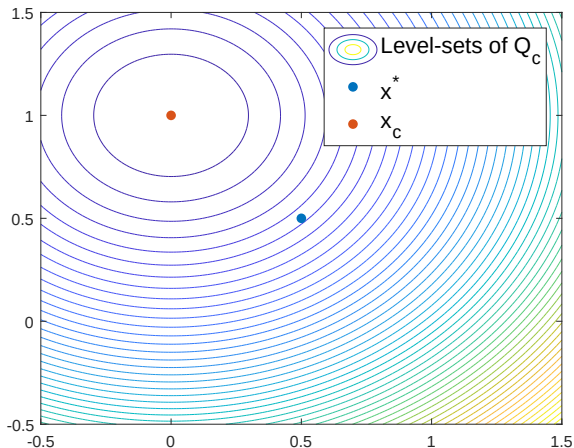


Figure: Graph of Q_c , for $c = 0$.

Quadratic penalization

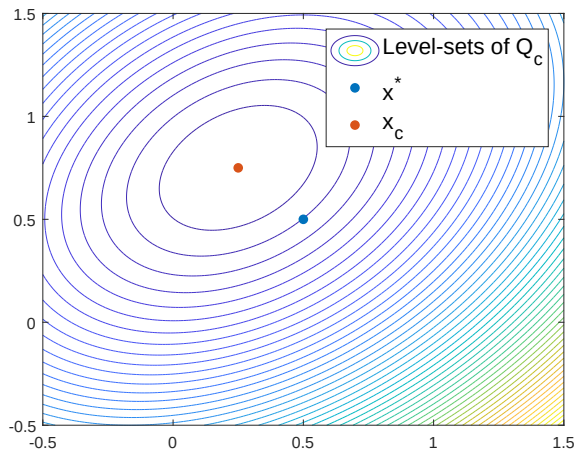


Figure: Graph of Q_c , for $c = 0.5$.

Quadratic penalization

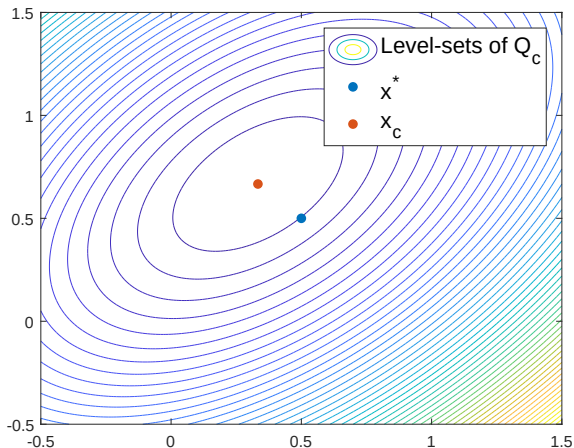


Figure: Graph of Q_c , for $c = 1$.

Quadratic penalization

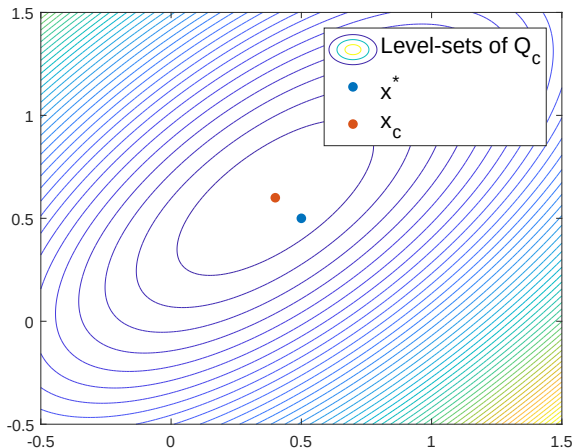


Figure: Graph of Q_c , for $c = 2$.

Quadratic penalization

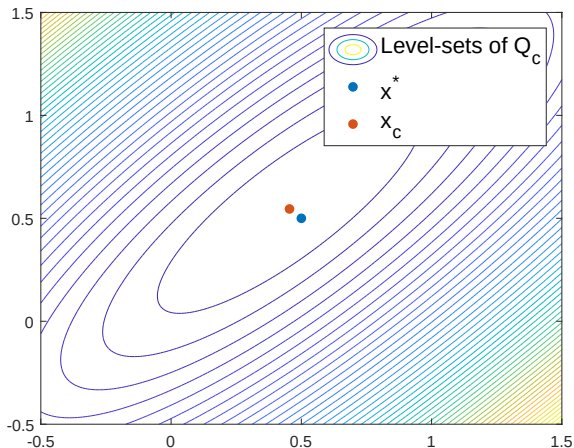


Figure: Graph of Q_c , for $c = 5$.

General quadratic case

Consider

$$\min_{x \in \mathbb{R}^n} \frac{1}{2} x^\top H x - q^\top x, \quad A x = b, \quad H \succ 0.$$

With $h(x) = Ax - b$, the penalty objective is

$$Q_c(x) = \frac{1}{2} x^\top H x - q^\top x + \frac{c}{2} (Ax - b)^\top (Ax - b).$$

By the chain rule,

$$\nabla_x \left(\frac{1}{2} \|Ax - b\|^2 \right) = A^\top (Ax - b).$$

Consequently,

$$\nabla Q_c(x) = Hx - q + cA^\top (Ax - b).$$

The stationarity equation becomes

$$(H + cA^\top A)x_c = q + cA^\top b.$$

For $H \succ 0$ and $c > 0$, the matrix is **SPD**.

Direct KKT vs. quadratic penalty

Direct KKT formulation

$$\begin{pmatrix} H & A^\top \\ A & 0 \end{pmatrix} \begin{pmatrix} x \\ \mu \end{pmatrix} = \begin{pmatrix} q \\ b \end{pmatrix}$$

- Constraints enforced to solver accuracy.
- Multipliers computed directly.
- Symmetric indefinite system: requires suitable solvers (e.g. MINRES, pivoted $LDL^\top$).

Quadratic penalization

$$(H + cA^\top A)x_c = q + cA^\top b$$

- SPD if $H \succ 0$; CG and Cholesky apply.
- Smaller primal-only system.
- Approximate feasibility for finite c .
- Conditioning may deteriorate as c grows.

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 $k = 0, \dots, K - 1$, do
 - Solve (P_{c_k}) (e.g. using gradient descent starting from x_k) and denote its solution by x_{k+1} .
 - If $\|h(x_{k+1})\|$ is below the prescribed tolerance, stop.
 - Otherwise, choose $c_{k+1} > c_k$.

End for.

- 3 Output: the last computed iterate.

Penalization

Let $c > 0$. The quadratic penalization $Q_c : \mathbb{R}^n \times \mathbb{R}^m \rightarrow \mathbb{R}$ is defined by

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

Its gradient with respect to x is

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

If $\bar{\mu}$ denotes the optimal Lagrange multiplier, we seek

$$c_k h(x_{k+1}) \simeq \bar{\mu},$$

or equivalently,

$$h(x_{k+1}) \simeq \frac{\bar{\mu}}{c_k}.$$

Augmented Lagrangian: multiplier update

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

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

Its gradient with respect to x is

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

If $\bar{\mu}$ denotes the optimal Lagrange multiplier, we seek

$$\mu_k + c_k h(x_{k+1}) \simeq \bar{\mu},$$

or equivalently,

$$h(x_{k+1}) \simeq \frac{\bar{\mu} - \mu_k}{c_k}.$$

This motivates the **multiplier update**:

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

Augmented Lagrangian: why?

Pure quadratic penalties use

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

A very large c enforces feasibility, but may badly condition the minimization subproblem. **Idea:** use an estimate of the Lagrange multiplier to achieve accurate feasibility *without forcing* $c \rightarrow \infty$.

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

- x is the primal variable; μ is the multiplier **estimate**.
- The method alternates a minimization in x and a multiplier update.

Augmented Lagrangian: multiplier update

After approximately minimizing $L_{c_k}(\cdot, \mu_k)$,

$$\begin{aligned} x_{k+1} &\approx \arg \min_x L_{c_k}(x, \mu_k), \\ \mu_{k+1} &= \mu_k + c_k h(x_{k+1}). \end{aligned}$$

Thus the new multiplier enforces the **stationarity relation**
 $\nabla_x L(x_{k+1}, \mu_{k+1}) \approx 0$.

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

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

- Augmented lagrangian:

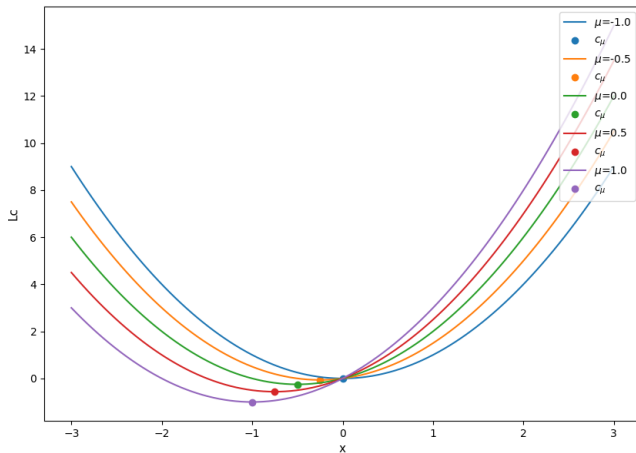
$$L_c(x, \mu) = x - x^2 + \mu x + \frac{c}{2}x^2 = (1 + \mu)x + \left(\frac{c}{2} - 1\right)x^2.$$

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

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

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

Augmented Lagrangian



Quadratic penalization

Example 2. Consider:

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

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

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 $c > 0$,

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

Solution.

- $\bar{y} - 1 + \bar{\mu} = 0$. We find $\bar{\mu} = 0.5$.

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 $\frac{1}{2}(x^2 + (y - 1)^2) + \frac{\varepsilon}{2}(x - y)^2 + \mu(x - y)$) 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 $1 - y$:

$$y - 1 + c(y - 1 + y) + \mu = y(1 + 2c) - c - 1 + \mu = 0, \text{ and thus,}$$

$$y = \frac{1+c-\mu}{1+2c}. \text{ So, } \begin{pmatrix} x_c \\ y_c \end{pmatrix} = \frac{1}{1+2c} \begin{pmatrix} c+\mu \\ 1+c-\mu \end{pmatrix}.$$

Quadratic penalization

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

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

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

Augmented Lagrangian

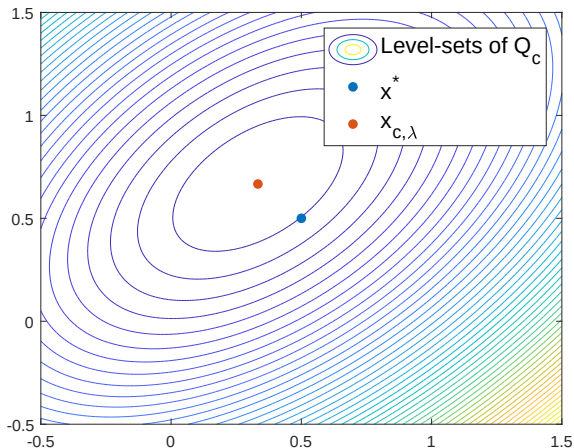


Figure: Level-sets $L_c(\cdot, \mu)$, for $c = 1$ and $\mu = 0$.

Augmented Lagrangian

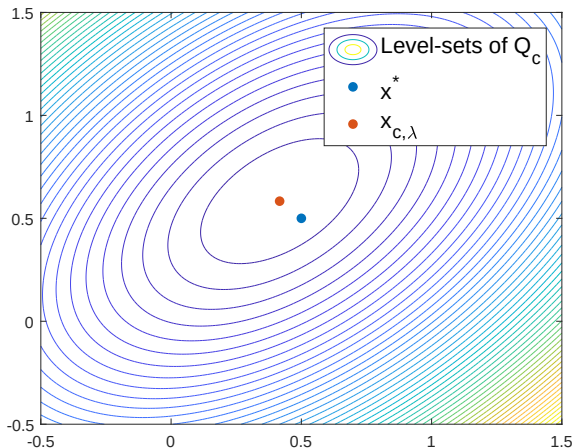


Figure: Level-sets $L_c(\cdot, \mu)$, for $c = 1$ and $\mu = 0, 25$.

Augmented Lagrangian

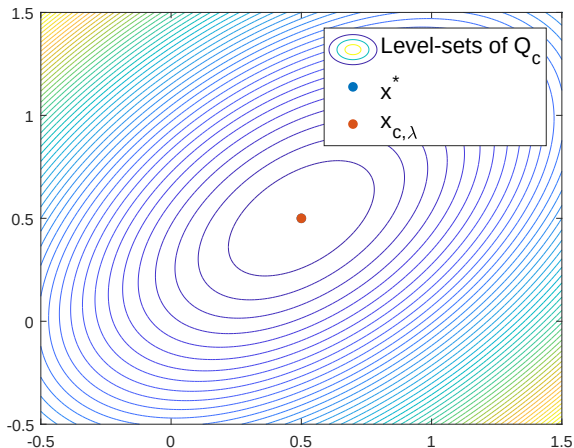


Figure: Level-sets $L_c(\cdot, \mu)$, for $c = 1$ and $\mu = 0, 5$.

Quadratic penalization

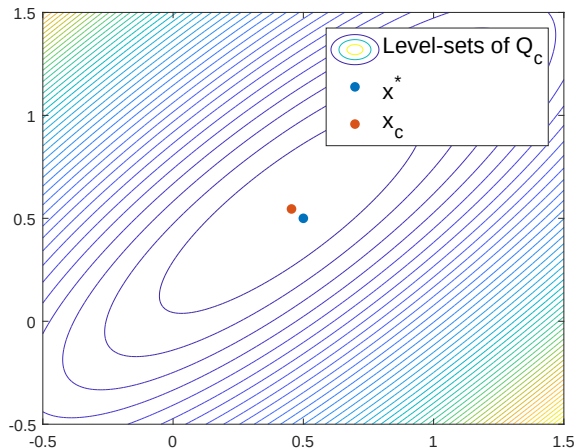


Figure: Graph of Q_c , for $c = 5$.

Augmented Lagrangian: algorithm

- 1 Choose $x_0 \in \mathbb{R}^n$, $\mu_0 \in \mathbb{R}^m$, $c_0 > 0$ and tolerances $\varepsilon_{\text{stat}}, \varepsilon_{\text{feas}} > 0$.
- 2 For $k = 0, 1, \dots$:
 - 1 Approximately minimize $L_{c_k}(x, \mu_k)$ in x to obtain x_{k+1} .
 - 2 Update $\mu_{k+1} = \mu_k + c_k h(x_{k+1})$.
 - 3 Stop when $\|\nabla_x L(x_{k+1}, \mu_{k+1})\| \leq \varepsilon_{\text{stat}}$ and $\|h(x_{k+1})\| \leq \varepsilon_{\text{feas}}$.
 - 4 Keep $c_{k+1} = c_k$ if feasibility improves sufficiently; otherwise increase $c_{k+1} > c_k$.

Practical advantage: multiplier updates improve feasibility without necessarily taking a very large penalty parameter.

Three approaches: take-home message

	Strength	Main difficulty
Direct KKT	Exact constraints and multipliers (to solver accuracy)	Indefinite saddle-point systems; nonlinear active sets
Quadratic penalty	Unconstrained subproblems; often SPD and sparse	Large c gives ill-conditioning; approximate feasibility
Augmented Lagrangian	Updates multiplier and feasibility with moderate c	Inner minimizations and tuning/stopping criteria

Penalty and augmented-Lagrangian methods are alternatives to solving KKT

equations directly, not guarantees of an easier problem.

1 Reminders

- KKT
- Sensitivity analysis

2 Penalty methods

- Quadratic penalization
- Augmented Lagrangian

3 Projected gradient method

- Projection
- Method
- Combination with penalty methods

Projection

Idea: Apply steepest descent method but project the path onto the constraints. The projected gradient method uses a mapping called **projection** defined below.

Lemma 2

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

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

It is called **projection** of x_0 on K , and denoted $\text{Proj}_K(x_0)$.

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

Projection

Example 1: projection onto a box.

Let K be described by

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

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

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

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

Projection

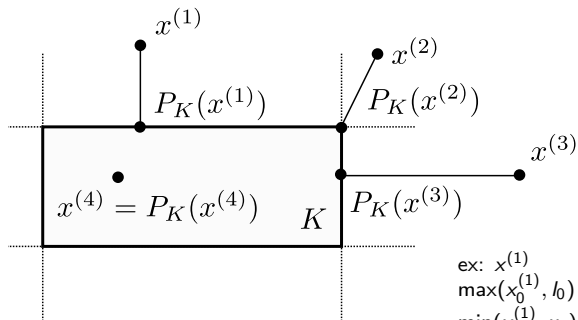


Figure: Projection on a cuboid.

$$\begin{aligned} \text{ex: } x^{(1)} \\ \max(x_0^{(1)}, l_0) &= x_0^{(1)} \\ \min(x_0^{(1)}, u_0) &= x_0^{(1)} \end{aligned}$$

$$\begin{aligned} \max(x_1^{(1)}, l_1) &= x_1^{(1)} \\ \min(x_1^{(1)}, u_1) &= u_1 \end{aligned}$$

Projection

Example 2: projection onto a ball.

Let K be described by

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

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

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

$$\text{Proj}_K(x) = \begin{cases} x_C, & x = x_C, \\ x_C + \min\left(1, \frac{R}{\|x - x_C\|}\right)(x - x_C), & x \neq x_C. \end{cases}$$

Projection

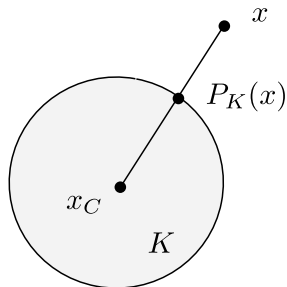


Figure: Projection on a ball.

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

$$\text{Proj}_K(x) = \left(\text{Proj}_{K_1}(x_1), \text{Proj}_{K_2}(x_2) \right).$$

Method

Optimization problem. Consider

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

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

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

Gradient descent algorithm.

- 1 Input: $x_0 \in \mathbb{R}^n$, $\varepsilon > 0$. Set $k = 0$.
- 2 While $\|\nabla f(x_k)\| \geq \varepsilon$, do
 - Find a descent direction d_k .
 - Find $\alpha_k > 0$ such that $f(x_k + \alpha_k d_k) < f(x_k)$.
 - Set $x_{k+1} = x_k + \alpha_k d_k$.
 - Set $k = k + 1$.

- 3 Output: x_k .

Main idea:

at iteration k , replace the search on the half line $\{x_k + \alpha_k d_k \mid \alpha \geq 0\}$ used in unconstrained optimization by a **search** on

$$\underbrace{\{\text{Proj}_K(x_k + \alpha_k d_k) \mid \alpha_k \geq 0\}}_{=: x_{k+1}(\alpha_k)}$$

Combination with penalty methods

Consider the problem

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

where $f: \mathbb{R}^n \rightarrow \mathbb{R}$, $h: \mathbb{R}^n \rightarrow \mathbb{R}^{m_1}$ and $g: \mathbb{R}^n \rightarrow \mathbb{R}^{m_2}$ are given.

Idea: Eliminate inequality constraints by slack variables. An equivalent formulation is

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

where: $\Phi_i(x) = \begin{cases} h_i(x), & \forall i \in \mathcal{E}, \\ g_i(x), & \forall i \in \mathcal{I}, \end{cases}$ and

$$K = \left\{ y \in \mathbb{R}^m \mid \begin{cases} y_i = 0 & \forall i \in \mathcal{E} \\ y_i \leq 0 & \forall i \in \mathcal{I} \end{cases} \right\}.$$

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

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

subject to: $y \in K$,

with the projected gradient method.

Use (x_k, y_k) as a starting point.