Constrained optimization

V. Leclère (ENPC)

April 16th, 2021

Why should I bother to learn this stuff ?

- Most real problems have constraints that you have to deal with.
- This course give a snapshot of the tools available to you.
- $\Longrightarrow$ useful for
  - having an idea of what can be done when you have constraints

Constrained optimization problem

- In the previous courses we have developped algorithms for unconstrained optimization problem.
- We now want to sketch some methods to deal with the constrained problem

$$\begin{aligned} \underset{x\in\mathbb{R}^n}{\text{Min}} \quad & f(x) \\ \text{s.t.} \quad & x \in X \end{aligned}$$

- We are going to discuss multiple type of constraint set X:
  - X is a ball : $\{x \mid \|x - x_0\|_2 \leq r\}$
  - X is a box : $\{x \mid \underline{x}_i \leq x_i \leq \bar{x}_i \quad \forall i \in [n]\}$
  - X is a polyhedron: $\{x \mid Ax \leq b\}$
  - X is given through explicit constraints $\{x \mid g(x) = 0, \quad h(x) \leq 0\}$

Contents

Contents

Admissible descent direction

- Recall that a descent direction d at point $x^{(k)} \in \mathbb{R}^n$ is a vector such that $\nabla f(x^{(k)})^\top d < 0$.
- An admissible descent direction at point $x^{(k)} \in X$ is a descent direction d such that, there exists $\varepsilon > 0$, such that, forall $t \leq \varepsilon$, $x^{(k)} + td \in X$.
- In other words, an admissible descent direction, is a direction that locally decrease the objective while staying in the constraint set.
- An admissible descent direction algorithm is naturally defined by:
  - A choice of admissible descent direction $d^{(k)}$
  - A choice of (sufficiently small) step $t^{(k)}$
  - $x^{(k+1)} = x^{(k)} + t^{(k)} d^{(k)} \in X$
- Warning : this does-not necessarilly converges. We can construct example where the step size get increasingly small because of the constraints.

A counter example

Consider

$$\begin{aligned} \min_{x \in \mathbb{R}^3} \quad & f(x) := \frac{4}{3}(x_1^2 - x_1 x_2 + x_2^2)^{3/4} - x_3 \\ \text{s.t.} \quad & x \geq 0 \end{aligned}$$

We set $x^{(0)} = (0, 2^{-3/2}, 0)$, and $d^{(k)}$ such that $d_i^{(k)} = -g_i^{(k)} \mathbb{1}_{x_i^{(k)} > 0}$, with $g_i^{(k)} = \nabla f(x^{(k)})$, and choose $t^{(k)}$ as the optimal step.

- This is an admissible direction descent with optimal step.
- f is strictly convex.
- $x^{(k)}$ converges toward a non-optimal point.

A counter example

Consider

$$\begin{aligned} \min_{x\in\mathbb{R}^3} \quad & f(x) := \frac{4}{3}(x_1^2 - x_1x_2 + x_2^2)^{3/4} - x_3 \\ \text{s.t.} \quad & x \geq 0 \end{aligned}$$

We set $x^{(0)} = (0, 2^{-3/2}, 0)$, and $d^{(k)}$ such that $d_i^{(k)} = -g_i^{(k)} \mathbb{1}_{x_i^{(k)}>0}$, with $g_i^{(k)} = \nabla f(x^{(k)})$, and choose $t^{(k)}$ as the optimal step.

- This is an admissible direction descent with optimal step.
- f is strictly convex.
- $x^{(k)}$ converges toward a non-optimal point.

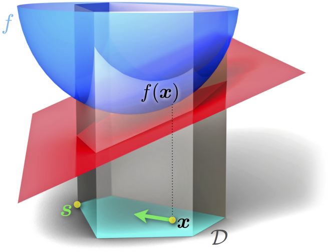
Conditional gradient algorithm

We address an optimization problem with convex objective function f and compact polyhedral constraint set X, i.e.

$$\min_{x \in X \subset \mathbb{R}^n} \qquad f(x)$$

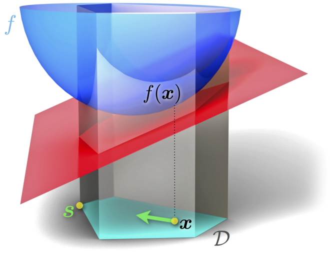
where

$$X = \{x \in \mathbb{R}^n \quad | \quad Ax \leq b, \quad \tilde{A}x = \tilde{b}\}$$



Conditional gradient algorithm

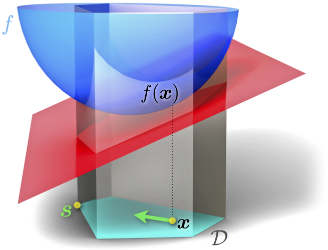
It is a descent algorithm, where we first look for an admissible descent direction $d^{(k)}$, and then look for the optimal step.



Conditional gradient algorithm

It is a descent algorithm, where we first look for an admissible descent direction $d^{(k)}$, and then look for the optimal step.
As f is convex, we know that for any point $x^{(k)}$,

$$f(y) \geq f(x^{(k)}) + \nabla f(x^{(k)}) \cdot (y - x^{(k)})$$

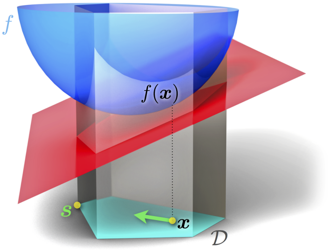


Conditional gradient algorithm

It is a descent algorithm, where we first look for an admissible descent direction $d^{(k)}$, and then look for the optimal step.
As f is convex, we know that for any point $x^{(k)}$,

$$f(y) \geq f(x^{(k)}) + \nabla f(x^{(k)}) \cdot (y - x^{(k)})$$

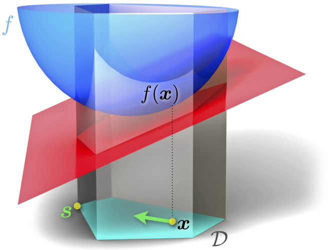
The conditional gradient method consists in choosing the descent direction that minimize the linearization of f over X.



Conditional gradient algorithm

The conditional gradient method consists in choosing the descent direction that minimize the linearization of f over X. More precisely, at step k we solve

$$y^{(k)} \in \underset{y \in X}{\arg\min} \quad f(x^{(k)}) + \nabla f(x^{(k)}) \cdot (y - x^{(k)}$$



Remarks on conditional gradient

$$y^{(k)} \in \arg\min_{y \in X} \quad f(x^{(k)}) + \nabla f(x^{(k)}) \cdot (y - x^{(k)}).$$

- This problem is linear, hence easy to solve.
- By the convexity inequality, the value of the linearized Problem is a lower bound to the true problem.
- As $y^{(k)} \in X$, $d^{(k)} = y^{(k)} - x^{(k)}$ is a *feasable direction*, in the sense that for all $t \in [0, 1]$, $x^{(k)} + td^{(k)} \in X$.
- If $y^{(k)}$ is obtained through the simplex method it is an extreme point of X, which means that, for $t > 1$, $x^{(k)} + td^{(k)} \notin X$.
- If $y^{(k)} = x^{(k)}$ then we have found an optimal solution.
- We also have $y^{(k)} \in \arg\min_{x \in X} \nabla f(x^{(k)}) \cdot y$, the lower-bound being obtained easily.

Remarks on conditional gradient

$$y^{(k)} \in \arg\min_{y \in X} \quad f(x^{(k)}) + \nabla f(x^{(k)}) \cdot (y - x^{(k)}).$$

- This problem is linear, hence easy to solve.
- By the convexity inequality, the value of the linearized Problem is a lower bound to the true problem.
- As $y^{(k)} \in X$, $d^{(k)} = y^{(k)} - x^{(k)}$ is a *feasable direction*, in the sense that for all $t \in [0,1]$, $x^{(k)} + td^{(k)} \in X$.
- If $y^{(k)}$ is obtained through the simplex method it is an extreme point of X, which means that, for $t > 1$, $x^{(k)} + td^{(k)} \notin X$.
- If $y^{(k)} = x^{(k)}$ then we have found an optimal solution.
- We also have $y^{(k)} \in \arg\min_{x \in X} \nabla f(x^{(k)}) \cdot y$, the lower-bound being obtained easily.

Remarks on conditional gradient

$$y^{(k)} \in \arg\min_{y \in X} \quad f(x^{(k)}) + \nabla f(x^{(k)}) \cdot (y - x^{(k)}).$$

- This problem is linear, hence easy to solve.
- By the convexity inequality, the value of the linearized Problem is a lower bound to the true problem.
- As $y^{(k)} \in X$, $d^{(k)} = y^{(k)} - x^{(k)}$ is a *feasable direction*, in the sense that for all $t \in [0,1]$, $x^{(k)} + td^{(k)} \in X$.
- If $y^{(k)}$ is obtained through the simplex method it is an extreme point of X, which means that, for $t > 1$, $x^{(k)} + td^{(k)} \notin X$.
- If $y^{(k)} = x^{(k)}$ then we have found an optimal solution.
- We also have $y^{(k)} \in \arg\min_{x \in X} \nabla f(x^{(k)}) \cdot y$, the lower-bound being obtained easily.

Remarks on conditional gradient

$$y^{(k)} \in \arg\min_{y \in X} \quad f(x^{(k)}) + \nabla f(x^{(k)}) \cdot (y - x^{(k)}).$$

- This problem is linear, hence easy to solve.
- By the convexity inequality, the value of the linearized Problem is a lower bound to the true problem.
- As $y^{(k)} \in X$, $d^{(k)} = y^{(k)} - x^{(k)}$ is a *feasable direction*, in the sense that for all $t \in [0,1]$, $x^{(k)} + td^{(k)} \in X$.
- If $y^{(k)}$ is obtained through the simplex method it is an extreme point of X, which means that, for $t > 1$, $x^{(k)} + td^{(k)} \notin X$.
- If $y^{(k)} = x^{(k)}$ then we have found an optimal solution.
- We also have $y^{(k)} \in \arg\min_{x \in X} \nabla f(x^{(k)}) \cdot y$, the lower-bound being obtained easily.

Remarks on conditional gradient

$$y^{(k)} \in \arg\min_{y \in X} \quad f(x^{(k)}) + \nabla f(x^{(k)}) \cdot (y - x^{(k)}).$$

- This problem is linear, hence easy to solve.
- By the convexity inequality, the value of the linearized Problem is a lower bound to the true problem.
- As $y^{(k)} \in X$, $d^{(k)} = y^{(k)} - x^{(k)}$ is a *feasable direction*, in the sense that for all $t \in [0,1]$, $x^{(k)} + td^{(k)} \in X$.
- If $y^{(k)}$ is obtained through the simplex method it is an extreme point of X, which means that, for $t > 1$, $x^{(k)} + td^{(k)} \notin X$.
- If $y^{(k)} = x^{(k)}$ then we have found an optimal solution.
- We also have $y^{(k)} \in \arg\min_{x \in X} \nabla f(x^{(k)}) \cdot y$, the lower-bound being obtained easily.

Remarks on conditional gradient

$$y^{(k)} \in \arg\min_{y \in X} \quad f(x^{(k)}) + \nabla f(x^{(k)}) \cdot (y - x^{(k)}).$$

- This problem is linear, hence easy to solve.
- By the convexity inequality, the value of the linearized Problem is a lower bound to the true problem.
- As $y^{(k)} \in X$, $d^{(k)} = y^{(k)} - x^{(k)}$ is a *feasable direction*, in the sense that for all $t \in [0, 1]$, $x^{(k)} + td^{(k)} \in X$.
- If $y^{(k)}$ is obtained through the simplex method it is an extreme point of X, which means that, for $t > 1$, $x^{(k)} + td^{(k)} \notin X$.
- If $y^{(k)} = x^{(k)}$ then we have found an optimal solution.
- We also have $y^{(k)} \in \arg\min_{x \in X} \nabla f(x^{(k)}) \cdot y$, the lower-bound being obtained easily.

Contents

Projection on a convex set ♡

Let $X \subset \mathbb{R}^n$ be a non-empty closed convex set. We call $P_X : \mathbb{R}^n \to \mathbb{R}^n$ the projection on X the fonction such that

$$P_X(x) = \underset{x' \in X}{\arg\min} \|x' - x\|_2^2$$

We have

- $\bar{x} = P_X(x)$ iff $(x - \bar{x}) \in N_X(\bar{x})$ (i.e. $\langle x - \bar{x}, x' - \bar{x} \rangle \leq 0, \quad \forall x' \in X$)
- $\langle P_X(y) - P_X(x), y - x \rangle \geq 0$ (P_X is *non-decreasing*)
- $\|P_X(y) - P_X(x)\|_2 \leq \|y - x\|$ (P_X is a *contraction*)

♠ Exercise: Prove these results

Projection on a convex set ♡

Let $X \subset \mathbb{R}^n$ be a non-empty closed convex set. We call $P_X : \mathbb{R}^n \to \mathbb{R}^n$ the projection on X the fonction such that

$$P_X(x) = \arg\min_{x' \in X} \|x' - x\|_2^2$$

We have

- $\bar{x} = P_X(x)$ iff $(x - \bar{x}) \in N_X(\bar{x})$ (i.e. $\langle x - \bar{x}, x' - \bar{x} \rangle \leq 0, \quad \forall x' \in X$)
- $\langle P_X(y) - P_X(x), y - x \rangle \geq 0$ (P_X is *non-decreasing*)
- $\|P_X(y) - P_X(x)\|_2 \leq \|y - x\|$ (P_X is a *contraction*)

♠ Exercise: Prove these results

Projected gradient

Consider

$$\begin{aligned} \underset{x \in \mathbb{R}^n}{\text{Min}} \quad & f(x) \\ \text{s.t.} \quad & x \in X \end{aligned}$$

where f is differentiable and X convex.

The projected gradient algorithm generate the following sequence

$$x^{(k+1)} = P_X\left[x^{(k)} - t^{(k)} g^{(k)}\right]$$

Projected gradient

II ◇

Theorem

Assume that $X \neq \emptyset$ is a closed convex set. $x^\sharp \in X$ is a critical point if and only if for one (or all) $t > 0$,

$$x^\sharp = P_X\big[x^\sharp - t\nabla f(x^\sharp)\big].$$

Theorem

If f is lower bounded on X, and with L-Lipschitz gradient, and X closed convex (non empty) set. Then the projected gradient algorithm with step staying in $[a, b] \subset]0, 2/L[$, then $\|x_{k+1} - x_k\| \to 0$, and any adherence point of $\{x_k\}_{k\in\mathbb{N}}$ is a critical point.

Corollary : if f convex differentiable with L-Lipschitz gradient, X compact convex non empty, the projected gradient algorithm with step $1/L$ is converging toward the optimal solution.

Theorem

Assume that $X \neq \emptyset$ is a closed convex set. $x^\sharp \in X$ is a critical point if and only if for one (or all) $t > 0$,

$$x^\sharp = P_X\big[x^\sharp - t\nabla f(x^\sharp)\big].$$

[redacted]

If f is lower bounded on X, and with L-Lipschitz gradient, and X closed convex (non empty) set. Then the projected gradient algorithm with step staying in $[a,b] \subset]0, 2/L[$, then $\|x_{k+1} - x_k\| \to 0$, and any adherence point of $\{x_k\}_{k\in\mathbb{N}}$ is a critical point.

Corollary : if f convex differentiable with L-Lipschitz gradient, X compact convex non empty, the projected gradient algorithm with step $1/L$ is converging toward the optimal solution.

When to use ? ♡

- Projected gradient is usefull only if the projection is simple, as projecting over a convex set consists in solving a constrained optimization problem.
- Projection is simple for balls and boxes.
- Finding an admissible direction is doable if the constraint set is polyhedral, or more generally conic-representable.

Contents

Contents

Idea of penalization

♡

We consider the constrained optimization problem

$$(\mathcal{P}) \quad \underset{x \in \mathbb{R}^n}{\text{Min}} \quad f(x)$$
$$\text{s.t.} \quad x \in X$$

and the following penalized version

$$(\mathcal{P}_t) \quad \underset{x \in \mathbb{R}^n}{\text{Min}} \quad f(x) + tp(x)$$

where $t > 0$, and $p : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ is a penalization function.

Thus, a (constrained) problem is replaced by a sequence of (unconstrained) problems.

♣ Exercise: What is happening if $p = \mathbb{I}_X$?

Idea of penalization ♡

We consider the constrained optimization problem

$$(\mathcal{P}) \quad \underset{x\in\mathbb{R}^n}{\text{Min}} \quad f(x)$$
$$\text{s.t.} \quad x \in X$$

and the following penalized version

$$(\mathcal{P}_t) \quad \underset{x\in\mathbb{R}^n}{\text{Min}} \quad f(x) + tp(x)$$

where $t > 0$, and $p : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ is a penalization function.

Thus, a (constrained) problem is replaced by a sequence of (unconstrained) problems.
♣ Exercise: What is happening if $p = \mathbb{I}_X$?

Some monotonicity results

$$(\mathcal{P}_t) \quad \underset{x \in \mathbb{R}^n}{\text{Min}} \qquad f(x) + tp(x)$$

The idea is that, with higher t, the penalization has more impact on the problem.
More precisely, let $0 < t_1 < t_2$, and x_{t_i} be an optimal solution of $(\mathcal{P}_{t_i})$.
We have:

- $p(x_{t_1}) \geq p(x_{t_2})$
- $f(x_{t_1}) \leq f(x_{t_2})$

♣ Exercise: prove these results.

Outer penalization

A first idea for choosing a penalization function p consists in choosing a function p such that:

- $p(x) = 0$ for $x \in X$
- $p(x) > 0$ for $x \notin X$

intuitively the idea is that p is the fine to pay for not respecting the constraint. Heuristically, it should be increasing with the distance to X.

Outer penalization - theoretical results

Assume that

- p is l.s.c on $\mathbb{R}^n$
- $p \geq 0$
- $p(x) = 0$ iff $x \in X$

Further assume that f is l.s.c and there exists $t_0 > 0$ such that $x \mapsto f(x) + t_0 p(x)$ is coercive (i.e. $\to \infty$ if $\|x\| \to \infty$).
Then,

1. for $t > t_0$, $(\mathcal{P}_t)$ admit at least one optimal solution
2. $(x_t)_{t\to+\infty}$ is bounded
3. any adherence point of $(x_t)_{t\to+\infty}$ is an optimal solution of $\mathcal{P}$.

Outer penalization - quadratic case

Assume that

$$X = \{x \in \mathbb{R}^n \quad | \quad g(x) = 0, \quad h(x) \leq 0\}$$

then the quadratic penalization consists in choosing

$$p : x \mapsto \|g(x)\|^2 + \|(h(x))^+\|^2$$

This choice is interesting as (for affinely lower-bounded f):

- $x \mapsto f(x) + \frac{1}{t}p(x)$ is differentiable if f is differentiable
- $x_t \to x^\sharp$ if $t \to 0$

However, generally speaking, if the constraints are impactful (e.g. have non-zero optimal multipliers), then

$$x_t \notin X$$

Outer penalization - quadratic case

Assume that

$$X = \left\{ x \in \mathbb{R}^n \quad | \quad g(x) = 0, \quad h(x) \leq 0 \right\}$$

then the quadratic penalization consists in choosing

$$p : x \mapsto \|g(x)\|^2 + \|(h(x))^+\|^2$$

This choice is interesting as (for affinely lower-bounded f):

- $x \mapsto f(x) + \frac{1}{t} p(x)$ is differentiable if f is differentiable
- $x_t \to x^\sharp$ if $t \to 0$

However, generally speaking, if the constraints are impactful (e.g. have non-zero optimal multipliers), then

$$x_t \notin X$$

Outer penalization - quadratic case

Assume that

$$X = \left\{ x \in \mathbb{R}^n \quad | \quad g(x) = 0, \quad h(x) \leq 0 \right\}$$

then the quadratic penalization consists in choosing

$$p : x \mapsto \|g(x)\|^2 + \|(h(x))^+\|^2$$

This choice is interesting as (for affinely lower-bounded f):

- $x \mapsto f(x) + \frac{1}{t}p(x)$ is differentiable if f is differentiable
- $x_t \to x^\sharp$ if $t \to 0$

However, generally speaking, if the constraints are impactful (e.g. have non-zero optimal multipliers), then

$$x_t \notin X$$

Outer penalization - L^1 case

Assume that

$$X = \left\{ x \in \mathbb{R}^n \quad | \quad g(x) = 0, \quad h(x) \leq 0 \right\}$$

another natural penalization consists in choosing

$$p : x \mapsto \|g(x)\|_1 + \|(h(x))^+\|_1$$

The interest of this approach is that, if the problem is convex and the constraints are qualified at optimality, then, for t small enough, an optimal solution to the penalized problem $(\mathcal{P}_t)$ is an optimal solution to the original problem $(\mathcal{P})$. Thus we speak of exact penalization.

Unfortunately this come to the price of non-differentiability.

Outer penalization - L^1 case

Assume that

$$X = \{x \in \mathbb{R}^n \quad | \quad g(x) = 0, \quad h(x) \leq 0\}$$

another natural penalization consists in choosing

$$p : x \mapsto \|g(x)\|_1 + \|(h(x))^+\|_1$$

The interest of this approach is that, if the problem is convex and the constraints are qualified at optimality, then, for t small enough, an optimal solution to the penalized problem $(\mathcal{P}_t)$ is an optimal solution to the original problem $(\mathcal{P})$. Thus we speak of exact penalization.

Unfortunately this come to the price of non-differentiability.

Outer penalization - L^1 case

Assume that

$$X = \{x \in \mathbb{R}^n \quad | \quad g(x) = 0, \quad h(x) \leq 0\}$$

another natural penalization consists in choosing

$$p : x \mapsto \|g(x)\|_1 + \|(h(x))^+\|_1$$

The interest of this approach is that, if the problem is convex and the constraints are qualified at optimality, then, for t small enough, an optimal solution to the penalized problem $(\mathcal{P}_t)$ is an optimal solution to the original problem $(\mathcal{P})$. Thus we speak of exact penalization.

Unfortunately this come to the price of non-differentiability.

Inner penalization

Another approach consists in choosing a penalization function that takes value $+\infty$ outside of X.

The idea here is to add a potential that repulse the optimal solution from the boundary.

This is typically done in a way to keep $f + tp$ smooth, and if possible convex.

More on that in the next course.

Inner penalization

Another approach consists in choosing a penalization function that takes value $+\infty$ outside of X.

The idea here is to add a potential that repulse the optimal solution from the boundary.

This is typically done in a way to keep $f + tp$ smooth, and if possible convex.

More on that in the next course.

Inner penalization

Another approach consists in choosing a penalization function that takes value $+\infty$ outside of X.

The idea here is to add a potential that repulse the optimal solution from the boundary.

This is typically done in a way to keep $f + tp$ smooth, and if possible convex.

More on that in the next course.

Contents

Duality, here we go again ♡

Recall that to a primal problem

$$(\mathcal{P}) \quad \underset{x\in\mathbb{R}^n}{\text{Min}} \quad f(x) \tag{1}$$
$$\text{s.t.} \quad g(x) = 0 \tag{2}$$
$$h(x) \leq 0 \tag{3}$$

we associate the dual problem

$$(\mathcal{D}) \quad \underset{\lambda,\mu\geq 0}{\text{Max}} \quad \underbrace{\underset{x}{\text{Min}} \quad f(x) + \lambda^\top g(x) + \mu^\top h(x)}_{\Phi(\lambda,\mu)}$$

♣ Exercise: Under which sufficient conditions are these problem equivalent ?

Duality, here we go again ♡

Recall that to a primal problem

$$(\mathcal{P}) \quad \underset{x\in\mathbb{R}^n}{\text{Min}} \quad f(x) \tag{1}$$
$$\text{s.t.} \quad g(x) = 0 \tag{2}$$
$$h(x) \leq 0 \tag{3}$$

we associate the dual problem

$$(\mathcal{D}) \quad \underset{\lambda,\mu\geq 0}{\text{Max}} \quad \underbrace{\underset{x}{\text{Min}} \quad f(x) + \lambda^\top g(x) + \mu^\top h(x)}_{\Phi(\lambda,\mu)}$$

♣ Exercise: Under which sufficient conditions are these problem equivalent ?

Duality seen as exact penalization ♡

If $(\mathcal{P})$ is convex differentiable and the constraints are qualified, then for any optimal multiplier $\overline{\lambda}, \overline{\mu}$ the unconstrained problem

$$\underset{x}{\text{Min}} \qquad f(x) + \overline{\lambda}^\top g(x) + \overline{\mu}^\top h(x)$$

have the same optimal solution as the original problem $(\mathcal{P})$.

Projected gradient in the dual

Consider the dual problem

$$(\mathcal{D}) \quad \underset{\lambda,\mu\geq 0}{\text{Max}} \qquad \Phi(\lambda,\mu)$$

Recall that, under technical conditions,

$$\nabla\Phi(\lambda,\mu) = \begin{pmatrix} g(x^\sharp(\lambda,\mu)) \\ h(x^\sharp(\lambda,\mu)) \end{pmatrix}$$

where $x^\sharp(\lambda,\mu)$ is an optimal solution of the inner minimization problem for given λ,μ.

We suggest to solve this problem through projected gradient with fixed step ρ:

$$\lambda^{(k+1)} = \lambda^{(k)} + \rho g(x^\sharp(\lambda^{(k)},\mu^{(k)}))$$
$$\mu^{(k+1)} = [\mu^{(k)} + \rho h(x^\sharp(\lambda^{(k)},\mu^{(k)}))]^+$$

Projected gradient in the dual

Consider the dual problem

$$(\mathcal{D}) \quad \underset{\lambda,\mu \geq 0}{\text{Max}} \qquad \Phi(\lambda, \mu)$$

Recall that, under technical conditions,

$$\nabla \Phi(\lambda, \mu) = \begin{pmatrix} g(x^\sharp(\lambda, \mu)) \\ h(x^\sharp(\lambda, \mu)) \end{pmatrix}$$

where $x^\sharp(\lambda, \mu)$ is an optimal solution of the inner minimization problem for given λ, μ.
We suggest to solve this problem through projected gradient with fixed step ρ:

$$\lambda^{(k+1)} = \lambda^{(k)} + \rho g(x^\sharp(\lambda^{(k)}, \mu^{(k)}))$$
$$\mu^{(k+1)} = [\mu^{(k)} + \rho h(x^\sharp(\lambda^{(k)}, \mu^{(k)}))]^+$$

Uzawa algorithm

Data: Initial primal point $x^{(0)}$, Initial dual points $\lambda^{(0)}, \mu^{(0)}$, unconstrained optimization method, dual step $\rho > 0$.

while $\|g(x^{(k)})\|_2 + \|(h(x^{(k)}))^+\|_2 \geq \varepsilon$ **do**

Solve for $x^{(k+1)}$

$$\underset{x}{\text{Min}} \qquad f(x) + \lambda^{(k)\top} g(x) + \mu^{(k)\top} h(x)$$

Update the multipliers

$$\lambda^{(k+1)} = \lambda^{(k)} + \rho g(x^{(k+1)})$$
$$\mu^{(k+1)} = [\mu^{(k)} + \rho h(x^{(k+1)})]^+$$

Algorithm 1: Uzawa algorithm

Convergence requires strong convexity and constraints qualifications.

Exercise : decomposition by prices

We consider the following energy problem:

- you are an energy producer with N production unit
- you have to satisfy a given demand planning for the next 24h (i.e. the total output at time t should be equal to d_t)
- the time step is the hour, and each unit have a production cost for each planning given as a convex quadratic function of the planning

1. Model this problem as an optimization problem. In which class does it belongs ? How many variables ?
2. Apply Uzawa's algorithm to this problem. Why could this be an interesting idea ?
3. Give an economic interpretation to this method.
4. What would happen if each unit had production constraints ?

What you have to know

- There is three main ways of dealing with constraints:
  - choosing an admissible direction
  - projection of the next iterate
  - penalizing the constraints

What you really should know

- admissible direction methods are mainly usefull for polyhedral constraint set
- projection is usefull only if the admissible set is simple (ball or bound constraints)
- penalization can be inner or outer, differentiable or not.

What you have to be able to do

- Implement a penalization approach.

What you should be able to do

- Implement Uzawa's algorithm.