CentraleSupelec
ST7 – Optimization
Part VIII: Some iterative algorithms

jean-christophe@pesquet.eu

迭代算法.

Problem

Let $\mathcal{H}$ be a Hilbert space.
Let $f : \mathcal{H} \to \mathbb{R}$ be Gâteaux differentiable.
Let C be a nonempty closed convex subset of $\mathcal{H}$.
We want to

$$\text{Find} \quad \widehat{x} \in \underset{x \in C}{\mathrm{Argmin}} \; f(x).$$

C. 非空闭凸子集、

Objective: Build a sequence $(x_n)_{n\in\mathbb{N}}$ converging to a minimizer.

收敛性.

Principle of first-order methods

- If f is Fréchet differentiable, then, at iteration n, we have

$$(\forall x \in \mathcal{H}) \qquad f(x) = f(x_n) + \langle \nabla f(x_n) \mid x - x_n \rangle + o(\|x - x_n\|).$$

  So if $\|x_{n+1} - x_n\|$ is small enough and x_{n+1} is chosen such that

$$\boxed{\langle \nabla f(x_n) \mid x_{n+1} - x_n \rangle < 0}$$

  then $f(x_{n+1}) < f(x_n)$.
- In particular, the steepest descent direction is given by

$$x_{n+1} - x_n = -\gamma_n \nabla f(x_n), \qquad \gamma_n \in \left]0, +\infty\right[.$$

learning rate.

Principle of first-order methods

- If f is Fréchet differentiable, then, at iteration n, we have

$$(\forall x \in \mathcal{H}) \qquad f(x) = f(x_n) + \langle \nabla f(x_n) \mid x - x_n \rangle + o(\|x - x_n\|).$$

  So if $\|x_{n+1} - x_n\|$ is small enough and x_{n+1} is chosen such that

$$\boxed{\langle \nabla f(x_n) \mid x_{n+1} - x_n \rangle < 0}$$

  then $f(x_{n+1}) < f(x_n)$.
- In particular, the steepest descent direction is given by

$$x_{n+1} - x_n = -\gamma_n \nabla f(x_n), \qquad \gamma_n \in \left]0, +\infty\right[.$$

- To secure that the solution belongs to C we can add a projection step.
- A relaxation parameter λ_n can also be added.

（松弛变量.

Projected gradient algorithm

The gradient algorithm has the following form:

$$(\forall n \in \mathbb{N}) \quad x_{n+1} = x_n - \gamma_n \nabla f(x_n)$$

where $\gamma_n \in \,]0, +\infty[$ is the stepsize. 步长?

Projected gradient algorithm

The projected gradient algorithm has the following form:

$$(\forall n \in \mathbb{N}) \quad x_{n+1} = P_C(x_n - \gamma_n \nabla f(x_n))$$

projection.

where $\gamma_n \in \,]0, +\infty[$.

Projected gradient algorithm

The projected gradient algorithm has the following form:

$$(\forall n \in \mathbb{N}) \quad x_{n+1} = x_n + \lambda_n\big(P_C(x_n - \gamma_n \nabla f(x_n)) - x_n\big)$$

where $\gamma_n \in \,]0, +\infty[$ and $\lambda_n \in]0, 1]$. relaxation.

Remark: x is a fixed point of the projected gradient iteration if and only if $x \in C$ and

$$(\forall y \in C) \qquad \langle \nabla f(x) \mid y - x \rangle \geq 0.$$

不动点迭代：

下一步迭代只与当前步的迭代点有关，

与其他迭代点无关.

Projected gradient algorithm

The projected gradient algorithm has the following form:

$$(\forall n \in \mathbb{N}) \quad x_{n+1} = x_n + \lambda_n\big(P_C(x_n - \gamma_n \nabla f(x_n)) - x_n\big)$$

where $\gamma_n \in \,]0, +\infty[$ and $\lambda_n \in]0, 1]$.

Proof: If x is a fixed point, then 不动点、

$$\begin{aligned} & x = x + \lambda_n\big(P_C(x - \gamma_n \nabla f(x)) - x\big) \\ \Leftrightarrow \quad & x = P_C(x - \gamma_n \nabla f(x)). \end{aligned}$$

According to the characterization of the projection, for every $y \in C$, ?

$$\begin{aligned} & \langle x - \gamma_n \nabla f(x) - x \mid y - x \rangle \leq 0 \\ \Leftrightarrow \quad & \langle \nabla f(x) \mid y - x \rangle \geq 0. \end{aligned}$$

投影的基本性质.

简单约束的凸优化问题 这一讲讨论简单约束可微凸优化问题

$$\min \{f(x) \mid x \in \Omega\}$$

的梯度算法, 其中 Ω 是$\Re^n$ 中的凸闭集, 并假设到 Ω 上的投影是容易实现的. 在第一讲中就已经提到, 简单约束可微凸优化问题等价于求变分不等式

$$\text{VI}(\Omega, \nabla f) \qquad x \in \Omega, \quad (x' - x)^T \nabla f(x) \geq 0, \quad \forall x' \in \Omega$$

的解. 这一讲的投影梯度方法, 分别是收缩算法和下降算法, 都不要用到函数值 $f(x)$, 只要对给定的 x, 能提供 $\nabla f(x)$. 收缩算法保证迭代点向解集靠近. 下降算法则隐含了目标函数值下降, 尽管目标函数值在计算过程中从不出现.

设 x^* 是变分不等式 VI$(\Omega, \nabla f)$ 的解. 由于 $\tilde{x} = P_\Omega[x - \beta\nabla f(x)] \in \Omega$, 因此根据变分不等式的定义有第一个基本不等式

$$\boxed{\text{(FI1)} \quad (\tilde{x} - x^*)^T \beta \nabla f(x^*) \geq 0.}$$

由于 $\tilde{x}$ 是 $x - \beta\nabla f(x)$ 在 Ω 上的投影, $x^* \in \Omega$, 根据投影的基本性质, 有

$$\boxed{\text{(FI2)} \quad (\tilde{x} - x^*)^T \big([x - \beta\nabla f(x)] - \tilde{x}\big) \geq 0.}$$

Projected gradient algorithm

The projected gradient algorithm has the following form:

$$(\forall n \in \mathbb{N}) \quad x_{n+1} = x_n + \lambda_n\big(P_C(x_n - \gamma_n \nabla f(x_n)) - x_n\big)$$

where $\gamma_n \in \,]0, +\infty[$ and $\lambda_n \in]0, 1]$.

Remark:

- x is a fixed point of the projected gradient iteration if and only if $x \in C$ and

凸的 $$(\forall y \in C) \qquad \langle \nabla f(x) \mid y - x \rangle \geq 0.$$

- When f is convex, x is a fixed point of the projected gradient iteration if and only if x is a global minimizer of f over C.

全局最小值.

Convergence in the convex case 收敛性.

Cocoercity property 强制性.

Assume that f is convex and has a Lipschtzian gradient with constant $\beta \in\,]0, +\infty[$, i.e.

$$(\forall (x,y) \in \mathcal{H}^2) \quad \|\nabla f(x) - \nabla f(y)\| \leq \beta \|x - y\|.$$

Then

$$(\forall (x,y) \in \mathcal{H}^2) \quad \beta \left\langle \nabla f(x) - \nabla f(y) \mid x - y \right\rangle \geq \|\nabla f(x) - \nabla f(y)\|^2.$$

Convergence in the convex case

Cocoercity property

Assume that f is convex and has a Lipschtzian gradient with constant $\beta \in \,]0, +\infty[$. Then

$$(\forall (x,y) \in \mathcal{H}^2) \quad \beta \langle \nabla f(x) - \nabla f(y) \mid x - y \rangle \geq \|\nabla f(x) - \nabla f(y)\|^2.$$

Convergence theorem

Assume that f is convex and has a Lipschtzian gradient with constant $\beta \in \,]0, +\infty[$.
Assume that $D = \mathrm{Argmin}_{x \in C} f(x) \neq \varnothing$.
Assume that $\inf_{n\in\mathbb{N}} \gamma_n > 0$, $\sup_{n\in\mathbb{N}} \gamma_n < 2/\beta$, $\inf_{n\in\mathbb{N}} \lambda_n > 0$, and $\sup_{n\in\mathbb{N}} \lambda_n \leq 1$.
Then the sequence $(x_n)_{n\in\mathbb{N}}$ generated by the projected gradient algorithm is Fejér monotone with respect to D, i.e.

$$(\forall x \in D)(\forall n \in \mathbb{N}) \quad \|x_{n+1} - x\| \leq \|x_n - x\|.$$

Convergence in the convex case

Convergence theorem

Assume that f is convex and has a Lipschtzian gradient with constant $\beta \in \,]0, +\infty[$.
Assume that $D = \mathrm{Argmin}_{x \in C} f(x) \neq \varnothing$.
Assume that $\inf_{n\in\mathbb{N}} \gamma_n > 0$, $\sup_{n\in\mathbb{N}} \gamma_n < 2/\beta$, $\inf_{n\in\mathbb{N}} \lambda_n > 0$, and $\sup_{n\in\mathbb{N}} \lambda_n \leq 1$.
Then the sequence $(x_n)_{n\in\mathbb{N}}$ generated by the projected gradient algorithm is Fejér monotone with respect to D.

Proof: Let $x \in D$. Then, for every $n \in \mathbb{N}$,

x is a fix point.

$x = P_C(x - \gamma_n \nabla f(x))$.

$$
\begin{aligned}
&\|P_C(x_n - \gamma_n \nabla f(x_n)) - x\|^2 \\
&= \|P_C(x_n - \gamma_n \nabla f(x_n)) - P_C(x - \gamma_n \nabla f(x))\|^2 \\
&\leq \|x_n - \gamma_n \nabla f(x_n) - x + \gamma_n \nabla f(x)\|^2 \quad \text{cocoercivity.} \\
&= \|x_n - x\|^2 - 2\gamma_n \langle \nabla f(x_n) - \nabla f(x) \mid x_n - x\rangle + \gamma_n^2 \|\nabla f(x_n) - \nabla f(x)\|^2 \\
&\leq \|x_n - x\|^2 - 2\gamma_n \beta^{-1} \|\nabla f(x_n) - \nabla f(x)\|^2 + \gamma_n^2 \|\nabla f(x_n) - \nabla f(x)\|^2 \\
&= \|x_n - x\|^2 - \gamma_n(2\beta^{-1} - \gamma_n)\|\nabla f(x_n) - \nabla f(x)\|^2 \leq \|x_n - x\|^2.
\end{aligned}
$$

Convergence in the convex case

Convergence theorem

Assume that f is convex and has a Lipschtzian gradient with constant $\beta \in]0,+\infty[$.
Assume that $D = \operatorname{Argmin}_{x\in C} f(x) \neq \varnothing$.
Assume that $\inf_{n\in\mathbb{N}} \gamma_n > 0$, $\sup_{n\in\mathbb{N}} \gamma_n < 2/\beta$, $\inf_{n\in\mathbb{N}} \lambda_n > 0$, and $\sup_{n\in\mathbb{N}} \lambda_n \leq 1$.
Then the sequence $(x_n)_{n\in\mathbb{N}}$ generated by the projected gradient algorithm is Fejér monotone with respect to D.

Proof: We deduce that

$$\begin{aligned}\|x_{n+1}-x\| &\leq (1-\lambda_n)\|x_n-x\| + \lambda_n\|P_C(x_n-\gamma_n\nabla f(x_n))-x\|\\ &\leq \|x_n-x\|.\end{aligned}$$

This shows that $(x_n)_{n\in\mathbb{N}}$ is Fejér monotone with respect to D.

Convergence in the convex case

Convergence theorem

Assume that f is convex and has a Lipschtzian gradient with constant $\beta \in]0, +\infty[$.

Assume that $\mathrm{Argmin}_{x \in C} f(x) \neq \varnothing$.

Assume that $\inf_{n \in \mathbb{N}} \gamma_n > 0$, $\sup_{n \in \mathbb{N}} \gamma_n < 2/\beta$, $\inf_{n \in \mathbb{N}} \lambda_n > 0$, and $\sup_{n \in \mathbb{N}} \lambda_n \leq 1$.

Then the sequence $(x_n)_{n \in \mathbb{N}}$ generated by the projected gradient algorithm converges weakly to a minimizer of f over C.

梯度投影法、

弱收敛 _

Convergence of the function values

Descent lemma

Assume that f is Gâteaux differentiable and has a β-Lipschtzian gradient with $\beta \in]0, +\infty[$. Then,

$$(\forall (x, y) \in \mathcal{H}^2) \quad f(y) \leq f(x) + \langle y - x \mid \nabla f(x) \rangle + \frac{\beta}{2} \|y - x\|^2.$$

Proof: For every $(x, y) \in \mathcal{H}^2$ and $t \in \mathbb{R}$, let $\varphi(t) = f(x + t(y - x))$.
φ is differentiable and $\varphi'(t) = \langle y - x \mid \nabla f(x + t(y - x)) \rangle$. We have then

$$\varphi(1) - \varphi(0) = \int_0^1 \varphi'(t) dt$$

$$\Leftrightarrow f(y) - f(x) - \langle y - x \mid \nabla f(x) \rangle = \int_0^1 \langle y - x \mid \nabla f(x + t(y - x)) - \nabla f(x) \rangle \, dt.$$

In addition, according to the Cauchy-Schwarz inequality,

$$\begin{aligned} &\langle y - x \mid \nabla f(x + t(y - x)) - \nabla f(x) \rangle \\ &\leq \|y - x\| \|\nabla f(x + t(y - x)) - \nabla f(x)\| \leq t\beta \|y - x\|^2. \end{aligned}$$

This leads to $f(y) \leq f(x) + \langle y - x \mid \nabla f(x) \rangle + \frac{\beta}{2} \|y - x\|^2$.

Convergence of the function values 函數值的收斂性

Descent lemma

Assume that f is Gâteaux differentiable and has a β-Lipschtzian gradient with $\beta \in \left]0,+\infty\right[$. Then,

$$(\forall (x,y) \in \mathcal{H}^2) \quad f(y) \leq f(x) + \langle y - x \mid \nabla f(x) \rangle + \frac{\beta}{2}\|y - x\|^2.$$

Convergence theorem

Assume that f is Gâteaux differentiable and has a β-Lipschtzian gradient with $\beta \in \left]0,+\infty\right[$. Assume that $\mu = \inf_{x \in C} f(x) > -\infty$.
Assume that $(\forall n \in \mathbb{N})$ $\lambda_n = 1$ and $\gamma_n \in]0, 1/\beta[$.
Then $\big(f(x_n)\big)_{n\in\mathbb{N}}$ is a convergent sequence.

Convergence of the function values

Convergence theorem

Assume that f is Gâteaux differentiable and has a β-Lipschtzian gradient with $\beta \in\ ]0,+\infty[$. Assume that $\mu = \inf_{x\in C} f(x) > -\infty$.
Assume that $(\forall n \in \mathbb{N})$ $\lambda_n = 1$ and $\gamma_n \in]0, 1/\beta[$.
Then $\big(f(x_n)\big)_{n\in\mathbb{N}}$ is a convergent sequence.

Proof: Let $n \geq 1$. According to the descent lemma,

$$f(x_{n+1}) \leq f(x_n) + \langle \nabla f(x_n) \mid x_{n+1} - x_n \rangle + \frac{\beta}{2}\|x_{n+1} - x_n\|^2$$

Since $x_{n+1} = P_C(x_n - \gamma_n \nabla f(x_n))$ and $x_n \in C$,

$$\begin{aligned}
& \|x_{n+1} - x_n + \gamma_n \nabla f(x_n)\|^2 \leq \|x_n - x_n + \gamma_n \nabla f(x_n)\|^2 \\
\Leftrightarrow\quad & \|x_{n+1} - x_n\|^2 + 2\gamma_n \langle \nabla f(x_n) \mid x_{n+1} - x_n \rangle \leq 0 \\
\Leftrightarrow\quad & \langle \nabla f(x_n) \mid x_{n+1} - x_n \rangle \leq -\frac{1}{2\gamma_n}\|x_{n+1} - x_n\|^2.
\end{aligned}$$

Therefore,

$$f(x_{n+1}) \leq f(x_n) + \frac{1}{2}(\beta - \gamma_n^{-1})\|x_{n+1} - x_n\|^2 \leq f(x_n).$$

Since $(f(x_n))_{n\in\mathbb{N}}$ is a decaying sequence, lower bounded by μ, it converges.

Convergence of the function values

Convergence theorem

Assume that f is Gâteaux differentiable and has a β-Lipschtzian gradient with $\beta \in \,]0, +\infty[$. Assume that $\mu = \inf_{x \in C} f(x) > -\infty$.
Assume that $(\forall n \in \mathbb{N})$ $\lambda_n = 1$ and $\gamma_n \in]0, 1/\beta[$.
Then $\big(f(x_n)\big)_{n \in \mathbb{N}}$ is a convergent sequence.

Remarks:

- If f is convex, the same result holds when $(\forall n \in \mathbb{N})$ $\lambda_n \in]0, 1]$ and $\gamma_n \in]0, 2/\beta[$. In addition, $f(x_n) \to \mu$.
- In the nonconvex case, there is no guarantee that the limit is μ since the iterates may get stuck in a spurious local minimum.

Metric change 度量变化.

Let $A \in \mathcal{B}(\mathcal{H}, \mathcal{H})$ be a self-adjoint operator which is strongly positive, i.e. there exists $\alpha \in \,]0, +\infty[$ such that

$$(\forall x \in \mathcal{H}) \qquad \langle x \mid Ax \rangle \geq \alpha \|x\|^2.$$

The inner product induced by A is

$$(\forall (x, y) \in \mathcal{H}^2) \qquad \langle x \mid y \rangle_A = \langle x \mid Ay \rangle \,.$$

Remark: When $\mathcal{H} = \mathbb{R}^N$, $A \in \mathbb{R}^{N \times N}$ is strongly positive if and only if A is symmetric positive definite.

Metric change

Let $A \in \mathcal{B}(\mathcal{H}, \mathcal{H})$ be a self-adjoint operator which is strongly positive, i.e. there exists $\alpha \in \,]0, +\infty[$ such that

$$(\forall x \in \mathcal{H}) \qquad \langle x \mid Ax \rangle \geq \alpha \|x\|^2.$$

The inner product induced by A is

$$(\forall (x, y) \in \mathcal{H}^2) \qquad \langle x \mid y \rangle_A = \langle x \mid Ay \rangle \,.$$

Let $A \in \mathcal{B}(\mathcal{H}, \mathcal{H})$ be a strongly positive self-adjoint operator.
Let $f : \mathcal{H} \to \mathbb{R}$ be a Gâteaux differentiable function.
In the Hilbert space $(\mathcal{H}, \langle \cdot \mid \cdot \rangle_A)$, the gradient of f is $\nabla_A f = A^{-1} \nabla f$.

Metric change

Let $A \in \mathcal{B}(\mathcal{H},\mathcal{H})$ be a self-adjoint operator which is strongly positive, i.e. there exists $\alpha \in \,]0,+\infty[$ such that

$$(\forall x \in \mathcal{H}) \qquad \langle x \mid Ax\rangle \geq \alpha \|x\|^2.$$

The inner product induced by A is

$$(\forall (x,y) \in \mathcal{H}^2) \qquad \langle x \mid y\rangle_A = \langle x \mid Ay\rangle\,.$$

Let $A \in \mathcal{B}(\mathcal{H},\mathcal{H})$ be a strongly positive self-adjoint operator.
Let $f\colon \mathcal{H} \to \mathbb{R}$ be a Gâteaux differentiable function.
In the Hilbert space $(\mathcal{H}, \langle\cdot \mid \cdot\rangle_A)$, the gradient of f is $\nabla_A f = A^{-1}\nabla f$.

Proof: The Gâteaux differential is such that

$$\boxed{\begin{aligned}(\forall (x,y) \in \mathcal{H}^2) \qquad f'(x)y &= \langle \nabla f(x) \mid y\rangle \\ &= \big\langle AA^{-1}\nabla f(x) \mid y\big\rangle \\ &= \big\langle A^{-1}\nabla f(x) \mid Ay\big\rangle \\ &= \big\langle \underline{A^{-1}\nabla f(x)} \mid y\big\rangle_A\,.\end{aligned}}$$

$\nabla_A f$.

Preconditioning

- Unconstrained optimization: $C = \mathcal{H}$ 无约束约化.
- Let $A \in \mathcal{B}(\mathcal{H}, \mathcal{H})$ be a strongly positive self-adjoint operator.
- The gradient algorithm in $(\mathcal{H}, \langle \cdot \mid \cdot \rangle_A)$ reads

$$(\forall n \in \mathbb{N}) \quad x_{n+1} = x_n - \gamma_n \nabla_A f(x_n)$$
$$= x_n - \gamma_n A^{-1} \nabla f(x_n)$$

with $\gamma_n \in \,]0, +\infty[$.

Preconditioning

- Unconstrained optimization: $C = \mathcal{H}$
- Let $A \in \mathcal{B}(\mathcal{H}, \mathcal{H})$ be a strongly positive self-adjoint operator.
- The gradient algorithm in $(\mathcal{H}, \langle \cdot \mid \cdot \rangle_A)$ reads

$$(\forall n \in \mathbb{N}) \quad \begin{aligned} x_{n+1} &= x_n - \gamma_n \nabla_A f(x_n) \\ &= x_n - \gamma_n A^{-1} \nabla f(x_n) \end{aligned}$$

  with $\gamma_n \in \,]0, +\infty[$.
- Let $(A_n)_{n \in \mathbb{N}}$ be a sequence of strongly positive self-adjoint operators in $\mathcal{B}(\mathcal{H}, \mathcal{H})$. A more general form is

$$\boxed{(\forall n \in \mathbb{N}) \quad \begin{aligned} x_{n+1} &= x_n - \gamma_n A_n^{-1} \nabla f(x_n) \\ &= x_n - \widetilde{A}_n^{-1} \nabla f(x_n), \end{aligned}}$$

  where $\widetilde{A}_n = \gamma_n^{-1} A_n$.
  $\rightsquigarrow$ quasi-Newton algorithm. 拟牛顿法.

Preconditioning

- Let $(A_n)_{n\in\mathbb{N}}$ be a sequence of strongly positive self-adjoint operators in $\mathcal{B}(\mathcal{H},\mathcal{H})$. A more general form is

$$\begin{aligned}(\forall n\in\mathbb{N})\quad x_{n+1} &= x_n - \gamma_n A_n^{-1}\nabla f(x_n)\\ &= x_n - \widetilde{A}_n^{-1}\nabla f(x_n),\end{aligned}$$

  where $\widetilde{A}_n = \gamma_n^{-1}A_n$.
  $\rightsquigarrow$ quasi-Newton algorithm .

<u>Remarks</u>:

- By being more flexible, this algorithm may lead to a faster convergence by a suitable choice of $(\widetilde{A}_n)_{n\in\mathbb{N}}$.
- If f is twice Fréchet differentiable and its Hessian is strongly positive on $\mathcal{H}$, one can choose

$$(\forall n\in\mathbb{N})\quad \widetilde{A}_n = \nabla^2 f(x_n)$$

  $\rightsquigarrow$ Newton's method . 牛顿法.

Preconditioning

Remarks:

- By being more flexible, this algorithm may lead to a faster convergence by a suitable choice of $(\widetilde{A}_n)_{n\in\mathbb{N}}$.
- If f is twice Fréchet differentiable and its Hessian is strongly positive on $\mathcal{H}$, one can choose

$$(\forall n \in \mathbb{N}) \quad \widetilde{A}_n = \nabla^2 f(x_n)$$

$\rightsquigarrow$ Newton's method.

- Newton's method can also be derived from a second-order Taylor expansion of f around x_n at iteration $n \in \mathbb{N}$: 二阶Taylor展开.

$f'(x_n)(x-x_n)$

$$x_{n+1} = \underset{x\in\mathcal{H}}{\operatorname{argmin}} \; f(x_n) + \langle \nabla f(x_n) \mid x - x_n \rangle + \frac{1}{2}\left\langle x - x_n \mid \nabla^2 f(x_n)(x - x_n)\right\rangle .$$

$f''(x_n)(x-x_n)^2/2$

Example 1: Uzawa algorithm

Problem

Let $\mathcal{L}\colon \mathcal{H} \times \mathbb{R}^q \to \mathbb{R}$ be differentiable with respect to its second argument. We want to find a saddle point of $\mathcal{L}$ over $\mathcal{H} \times [0,+\infty[^q$.

Solution

Set $\lambda_0 \in [0,+\infty[^q$

For $n = 0, 1, \ldots$

$$\left\lfloor \begin{array}{l} \text{Set } \gamma_n \in \;]0,+\infty[, \; \rho_n \in]0,1] \\ x_n \in \operatorname{Argmin} \mathcal{L}(\cdot, \lambda_n) \\ \lambda_{n+1} = \lambda_n + \rho_n \big(P_{[0,+\infty[^q}(\lambda_n + \gamma_n \nabla_\lambda \mathcal{L}(x_n, \lambda_n)) - \lambda_n\big). \end{array} \right.$$

Example 2: DC programming

Problem

Let $f \in \Gamma_0(\mathcal{H})$ and let $g \in \Gamma_0(\mathcal{H})$.
We want to minimize the difference of convex functions $f - g$.

Remark: The problem is equivalent to

$$
\begin{aligned}
& \underset{x \in \mathcal{H}}{\operatorname{minimize}} \; f(x) - \sup_{v \in \mathcal{H}} \left(\langle x \mid v \rangle - g^*(v) \right) \\
\Leftrightarrow \quad & \underset{(x,v) \in \mathcal{H}^2}{\operatorname{minimize}} \; f(x) - \langle x \mid v \rangle + g^*(v)
\end{aligned}
$$

Example 2: DC programming

Problem

Let $f \in \Gamma_0(\mathcal{H})$ and let $g \in \Gamma_0(\mathcal{H})$.
We want to minimize the difference of convex functions $f - g$.

Remark: The problem is equivalent to

$$\begin{aligned} &\underset{x\in\mathcal{H}}{\text{minimize}} \;\; f(x) - \sup_{v\in\mathcal{H}}\big(\langle x \mid v\rangle - g^*(v)\big) \\ \Leftrightarrow \quad &\underset{(x,v)\in\mathcal{H}^2}{\text{minimize}} \;\; f(x) - \langle x \mid v\rangle + g^*(v) \end{aligned}$$

Solution

If f and g^* are Gâteaux differentiable, we can use the following algorithm:

Set $(x_0, v_0) \in \mathcal{H}^2$
For $n = 0, 1, \dots$

$$\left\lfloor \begin{array}{l} \text{Set } \gamma_n \in\,]0,+\infty[,\ \mu_n \in\,]0,+\infty[\\ x_{n+1} = x_n - \gamma_n(\nabla f(x_n) - v_n) \\ v_{n+1} = v_n - \mu_n(\nabla g^*(v_n) - x_{n+1}). \end{array} \right.$$

Exercise

Let $\mathcal{H}$ and $\mathcal{G}$ be real Hilbert spaces and let $L \in \mathcal{B}(\mathcal{H}, \mathcal{G})$. Let $y \in \mathcal{G}$ and let $\alpha \in \,]0, +\infty[$.
We want to minimize the function defined as

$$(\forall x \in \mathcal{H}) \qquad f(x) = \frac{1}{2}\|Lx - y\|^2 + \frac{\alpha}{2}\|x\|^2.$$

1. Give the form of the gradient descent algorithm allowing us to solve this problem.
2. How does Newton's method read for this function ?
3. Consider the case when $\mathcal{H} = \mathbb{R}^N$. Study the convergence of the gradient descent algorithm by performing the eigendecomposition of $L^* L$.