Algorithmes pour l'optimisation semi-algébrique

Guillaume Garrigos

Institut de Mathématiques et de Modélisation de Montpellier
Universidad Tecnica Federico Santa Maria

Journées MODE Rennes
26 Mars 2014

Introduction : Problèmes semi-algébriques (1)

Soit $A \in \mathbb{R}^{M \times N}$ t.q. $M << N$. On veut résoudre $Ax = b$ dans R^N.

Soit $A \in \mathbb{R}^{M \times N}$ t.q. $M << N$. On veut résoudre $Ax = b$ dans R^N.
Définissons $\|x\|_0 = \#\{x_i \mid x_i \neq 0\}$ et considérons :

$$(P_1) \qquad \min_{x \in \mathbb{R}^N} \|x\|_0 + \delta_{Ax=b}(x)$$

Soit $A \in \mathbb{R}^{M\times N}$ t.q. $M << N$. On veut résoudre $Ax = b$ dans R^N.
Définissons $\|x\|_0 = \#\{x_i \mid x_i \neq 0\}$ et considérons :

$$(P_1) \qquad \min_{x\in\mathbb{R}^N} \|x\|_0 + \delta_{Ax=b}(x)$$

Peut être relaxé en :

$$(P_1') \qquad \min_{x\in\mathbb{R}^N} \alpha\|x\|_0 + \frac{1}{2}\|Ax - b\|^2$$

$$(P_1'') \qquad \min_{x,y\in\mathbb{R}^N} \alpha\|x\|_0 + \delta_{Ay=b}(y) + \frac{1}{2}\|x - y\|^2$$

Introduction : Problèmes semi-algébriques (2)

Soit $A \in \mathbb{R}^{M\times N}$. On veut une décomposition $A = X + Y$ sous la contrainte $\operatorname{rank} X \leq r$, $\|Y\|_0 \leq s$.

Soit $A \in \mathbb{R}^{M \times N}$. On veut une décomposition $A = X + Y$ sous la contrainte $\operatorname{rank} X \leq r$, $\|Y\|_0 \leq s$.

Revient à résoudre

$$(P_2) \qquad \min_{X,Y} \ \frac{1}{2}\|A - X - Y\|_F^2 + \delta_{\text{rank} \leq r}(X) + \delta_{\|\cdot\|_0 \leq s}(Y)$$

Introduction : Problèmes semi-algébriques

$$(P_1') \qquad \min_{x\in\mathbb{R}^N} \ \alpha\|x\|_0 + \frac{1}{2}\|Ax-b\|^2$$

$$(P_1'') \qquad \min_{x,y\in\mathbb{R}^N} \ \alpha\|x\|_0 + \delta_{Ay=b}(y) + \frac{1}{2}\|x-y\|^2$$

$$(P_2) \qquad \min_{X,Y} \ \frac{1}{2}\|A-X-Y\|_F^2 + \delta_{\text{rank} \leq r}(X) + \delta_{\|\cdot\|_0\leq s}(Y)$$

$\longrightarrow$ Problèmes non triviaux car non lisses non convexes (non continus !)
$\longrightarrow$ Problèmes structurés : parties lisses/non lisses.

Résultat de convergence

Soit $f : \mathbb{R}^N \longrightarrow \mathbb{R} \cup \{+\infty\}$ semi-algébrique.
Nous allons voir une classe d'algorithmes de décomposition qui vérifient :

Théorème (Frankel, G., Peypouquet, 2013 - Attouch et al., 2010)

1. Toute suite bornée est de longueur finie et convergente vers un point critique de f.
2. Si le point initial x_0 est suffisamment proche d'un minimum global a lors la suite est de longueur finie et converge vers un minimum global de f.

Résultat de convergence

Soit $f : \mathbb{R}^N \longrightarrow \mathbb{R} \cup \{+\infty\}$ semi-algébrique.
Nous allons voir une classe d'algorithmes de décomposition qui vérifient :

Théorème (Frankel, G., Peypouquet, 2013 - Attouch et al., 2010)

1. Toute suite bornée est de longueur finie et convergente vers un point critique de f.
2. Si le point initial x_0 est suffisamment proche d'un minimum global a lors la suite est de longueur finie et converge vers un minimum global de f.

$\rightarrow$ Se généralise en espaces de Hilbert
$\rightarrow$ Remplacer "semi-algébrique" par "Kurdyka-Lojasiewicz"

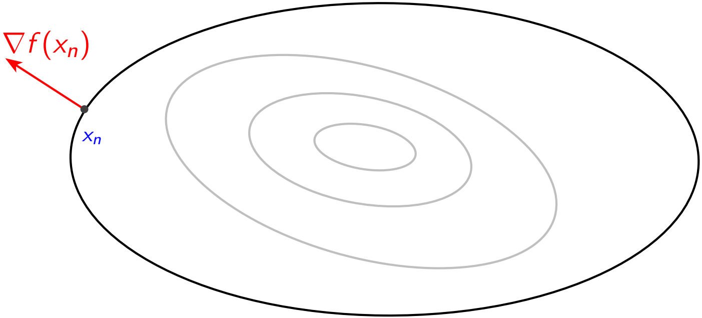
Sommaire

1. Descent methods
   - Explicit gradient method
   - Implicit gradient method

2. Convergence of splitting methods for semialgebraic functions

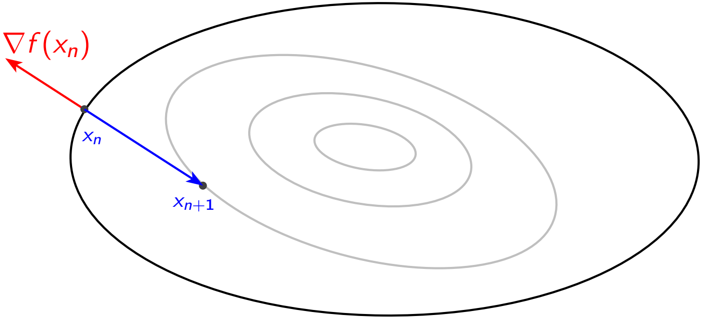
Méthode du gradient explicite (pas classique)

Méthode gradient explicite : $x_{n+1} = x_n - \lambda_n \nabla f(x_n)$



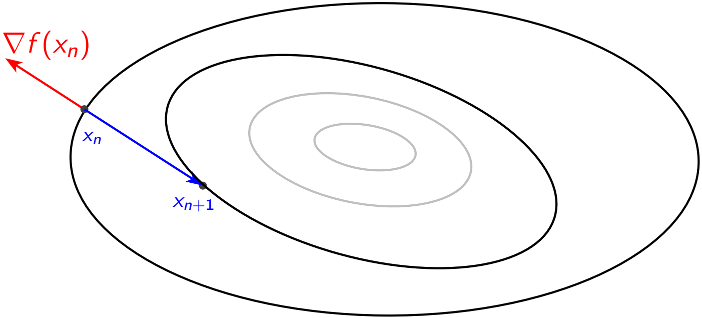
Méthode du gradient explicite (pas classique)

Méthode gradient explicite : $x_{n+1} = x_n - \lambda_n \nabla f(x_n)$



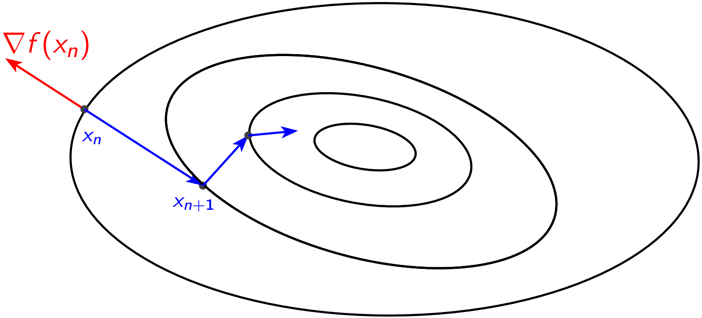
Méthode du gradient explicite (pas classique)

Méthode gradient explicite : $x_{n+1} = x_n - \lambda_n \nabla f(x_n)$



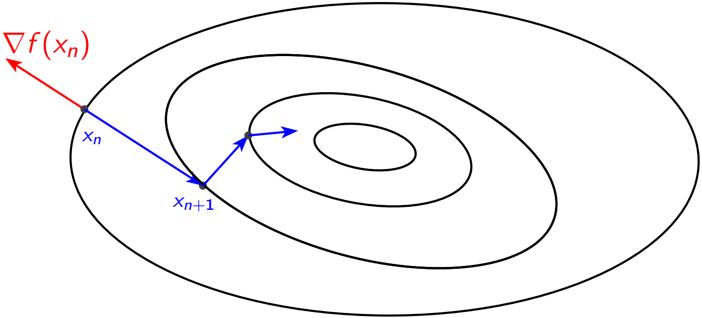
Méthode du gradient explicite (pas classique)

Méthode gradient explicite : $x_{n+1} = x_n - \lambda_n \nabla f(x_n)$



Méthode du gradient explicite (pas classique)

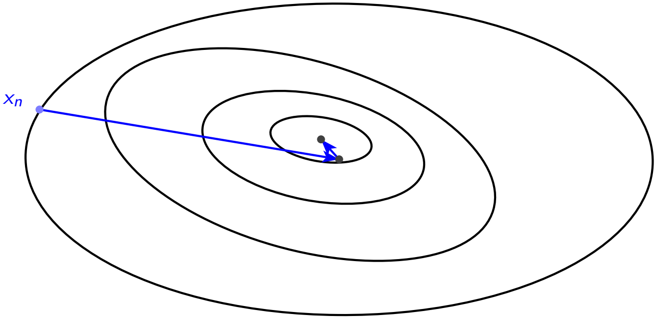
Méthode gradient explicite : $x_{n+1} = x_n - \lambda_n \nabla f(x_n)$



- Descente si ∇f est L-Lipschitz et $\lambda_n << \frac{2}{L}$
- Facile à calculer mais convergence lente. Corrigeable avec $\nabla^2 f(x_n)$ si DP.

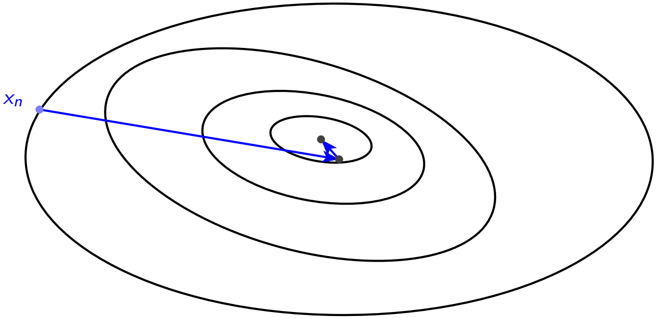
Méthode du gradient explicite (pas Newton)

Méthode de Newton : $x_{n+1} = x_n - \left(\nabla^2 f(x_n)\right)^{-1} \nabla f(x_n)$



Méthode du gradient explicite (pas Newton)

Méthode de Newton : $x_{n+1} = x_n - \left(\nabla^2 f(x_n)\right)^{-1} \nabla f(x_n)$



- Vitesse de convergence bien meilleure VS plus dur à calculer
- Ne marche que si $\nabla^2 f(x_n)$ existe et est inversible.

Méthode du gradient explicite (métrique variable)

Pour $A \in L(\mathbb{R}^N)$ symétrique on note

- $\alpha(A)$ l'inf des valeurs spectrales de A,
- $|||A|||$ le sup des valeurs spectrales en valeur absolue.

On considère $\mathcal{S}_{++}(\mathbb{R}^N)$ l'ensemble des opérateurs symétriques, définis positifs ($\alpha(A) > 0$) et bornés ($|||A||| < +\infty$).

Méthode du gradient explicite (métrique variable)

Pour $A \in L(\mathbb{R}^N)$ symétrique on note

- $\alpha(A)$ l'inf des valeurs spectrales de A,
- $|||A|||$ le sup des valeurs spectrales en valeur absolue.

On considère $\mathcal{S}_{++}(\mathbb{R}^N)$ l'ensemble des opérateurs symétriques, définis positifs ($\alpha(A) > 0$) et bornés ($|||A||| < +\infty$).

Pour $(A_n)_{n\in\mathbb{N}} \in \mathcal{S}_{++}(\mathbb{R}^N)$ on peut définir

$$x_{n+1} = x_n - A_n^{-1} \nabla f(x_n)$$

Méthode du gradient explicite (métrique variable)

Pour $A \in L(\mathbb{R}^N)$ symétrique on note

- $\alpha(A)$ l'inf des valeurs spectrales de A,
- $|||A|||$ le sup des valeurs spectrales en valeur absolue.

On considère $\mathcal{S}_{++}(\mathbb{R}^N)$ l'ensemble des opérateurs symétriques, définis positifs ($\alpha(A) > 0$) et bornés ($|||A||| < +\infty$).

Pour $(A_n)_{n\in\mathbb{N}} \in \mathcal{S}_{++}(\mathbb{R}^N)$ on peut définir

$$x_{n+1} = x_n - A_n^{-1}\nabla f(x_n)$$

- Avec $A_n = \frac{1}{\lambda_n} id$ on retrouve le gradient classique.
- Avec $A_n = \nabla^2 f(x_n) \in \mathcal{S}_{++}(\mathbb{R}^N)$ on retrouve Newton.
- Sinon on essaie $A_n \simeq \nabla^2 f(x_n)$ dans $\mathcal{S}_{++}(\mathbb{R}^N)$.

1. Descent methods
   - Explicit gradient method
   - Implicit gradient method

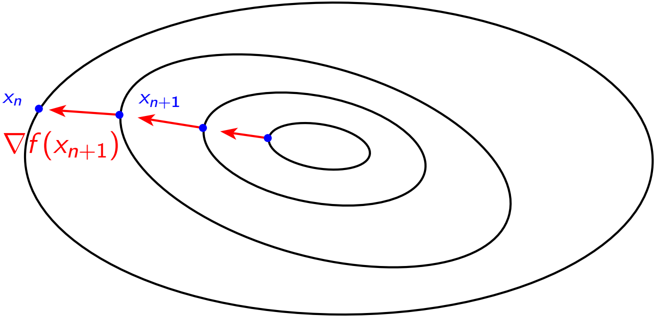
2. Convergence of splitting methods for semialgebraic functions

Méthode du gradient implicite (cas différentiable)

Méthode implicite : $x_{n+1} = x_n - \lambda_n \nabla f(x_{n+1})$

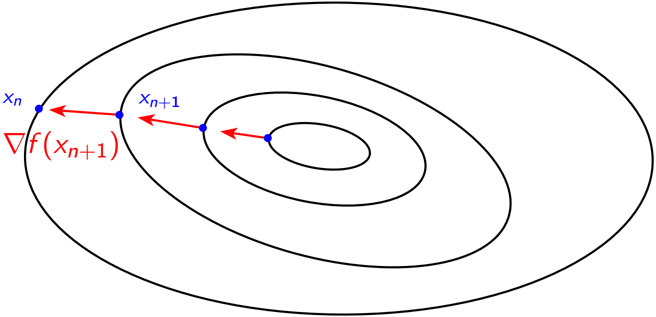
Méthode du gradient implicite (cas différentiable)

Méthode implicite : $x_{n+1} = x_n - \lambda_n \nabla f(x_{n+1})$



Méthode du gradient implicite (cas différentiable)

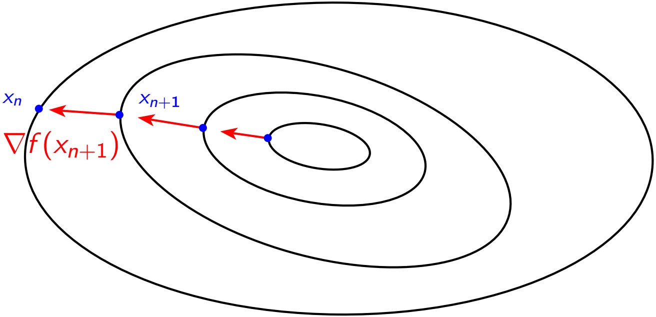
Méthode implicite : $x_{n+1} = x_n - \lambda_n \nabla f(x_{n+1})$



- Difficile à calculer mais convergence plus rapide + stabilité

Méthode du gradient implicite (cas différentiable)

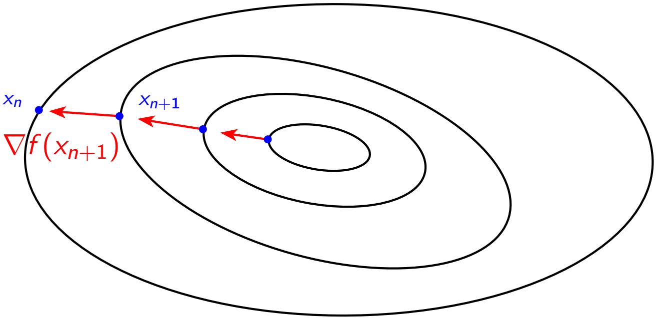
Méthode implicite : $x_{n+1} = x_n - A_n^{-1} \nabla f(x_{n+1})$



- Difficile à calculer mais convergence plus rapide + stabilité
- On peut aussi prendre en compte de la métrique variable

Méthode du gradient implicite (cas différentiable)

Méthode implicite : $x_{n+1} = x_n - A_n^{-1} \nabla f(x_{n+1})$



- Difficile à calculer mais convergence plus rapide + stabilité
- On peut aussi prendre en compte de la métrique variable
- S'adapte facilement aux fonctions non lisses

Supposons f convexe.

$$\begin{aligned} & x_{n+1} = x_n - A_n^{-1} \nabla f(x_{n+1}) \\ \Leftrightarrow \quad & A_n(x_{n+1} - x_n) + \nabla f(x_{n+1}) = 0 \end{aligned}$$

Algorithme du gradient implicite / proximal

Supposons f convexe. Posons $\|x\|_A^2 = \langle Ax, x\rangle$ pour $A \in \mathcal{S}_{++}(\mathbb{R}^N)$.

$$\begin{aligned} & x_{n+1} = x_n - A_n^{-1}\nabla f(x_{n+1}) \\ \Leftrightarrow \quad & A_n(x_{n+1} - x_n) + \nabla f(x_{n+1}) = 0 \\ \Leftrightarrow \quad & \nabla\left(\tfrac{1}{2}\|\cdot - x_n\|_{A_n}^2 + f(\cdot)\right)(x_{n+1}) = 0 \end{aligned}$$

Algorithme du gradient implicite / proximal

Supposons f convexe. Posons $\|x\|_A^2 = \langle Ax, x\rangle$ pour $A \in \mathcal{S}_{++}(\mathbb{R}^N)$.

$$\begin{array}{rc} & x_{n+1} = x_n - A_n^{-1}\nabla f(x_{n+1}) \\ \Leftrightarrow & A_n(x_{n+1} - x_n) + \nabla f(x_{n+1}) = 0 \\ \Leftrightarrow & \nabla\left(\frac{1}{2}\|\cdot - x_n\|_{A_n}^2 + f(\cdot)\right)(x_{n+1}) = 0 \\ \Leftrightarrow & x_{n+1} = Argmin_{y\in H}\left\{f(y) + \frac{1}{2}\|y - x_n\|_{A_n}^2\right\} \end{array}$$

Algorithme du gradient implicite / proximal

Supposons f convexe. Posons $\|x\|_A^2 = \langle Ax, x\rangle$ pour $A \in \mathcal{S}_{++}(\mathbb{R}^N)$.

$$\begin{array}{cc} & x_{n+1} = x_n - A_n^{-1}\nabla f(x_{n+1}) \\ \Leftrightarrow & A_n(x_{n+1} - x_n) + \nabla f(x_{n+1}) = 0 \\ \Leftrightarrow & \nabla\left(\frac{1}{2}\|\cdot - x_n\|_{A_n}^2 + f(\cdot)\right)(x_{n+1}) = 0 \\ \Leftrightarrow & x_{n+1} = Argmin_{y\in H}\left\{f(y) + \frac{1}{2}\|y - x_n\|_{A_n}^2\right\} \end{array}$$

Definition : Opérateur proximal

Soit $f : \mathbb{R}^N \to \mathbb{R} \cup \{+\infty\}$ s.c.i propre et $A \in \mathcal{S}_{++}(\mathbb{R}^N)$.

$$prox_f^A(x) := Argmin_{y\in H}\left\{f(y) + \frac{1}{2}\|y - x\|_A^2\right\}$$

On écrit alors $x_{n+1} \in \mathrm{prox}_f^{A_n}(x_n)$ et on parle d'algorithme proximal.

Definitions (Sous-différentiels)

$f : \mathbb{R}^N \to \mathbb{R} \cup \{+\infty\}$ s.c.i. propre. Soit $x \in \operatorname{dom} f$.

Sous-différentiel de Fréchet $\hat{\partial}$

$$p \in \hat{\partial} f(x) \Leftrightarrow \liminf_{y \to x,\ y \neq x} \frac{f(y) - f(x) - \langle p, y - x \rangle}{\|y - x\|} \geq 0$$

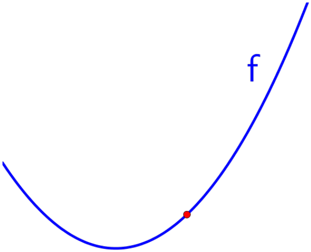
Definitions (Sous-différentiels)

$f : \mathbb{R}^N \to \mathbb{R} \cup \{+\infty\}$ s.c.i. propre. Soit $x \in \mathrm{dom}\, f$.

Sous-différentiel de Fréchet $\hat{\partial}$

$$p \in \hat{\partial} f(x) \Leftrightarrow \liminf_{y \to x,\ y \neq x} \frac{f(y) - f(x) - \langle p, y - x \rangle}{\|y - x\|} \geq 0$$

Si f est C^1



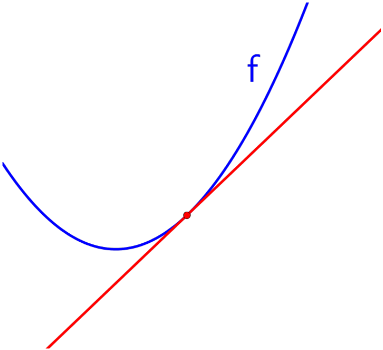
Definitions (Sous-différentiels)

$f : \mathbb{R}^N \to \mathbb{R} \cup \{+\infty\}$ s.c.i. propre. Soit $x \in \operatorname{dom} f$.

Sous-différentiel de Fréchet $\hat{\partial}$

$$p \in \hat{\partial} f(x) \Leftrightarrow \liminf_{y \to x,\ y \neq x} \frac{f(y) - f(x) - \langle p, y - x \rangle}{\|y - x\|} \geq 0$$

Si f est C^1 alors
$\hat{\partial} f(x) = \{\nabla f(x)\}$

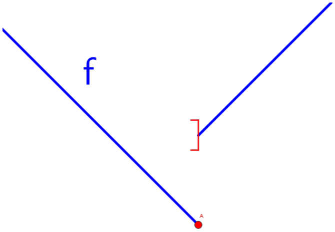


Definitions (Sous-différentiels)

$f : \mathbb{R}^N \to \mathbb{R} \cup \{+\infty\}$ s.c.i. propre. Soit $x \in \operatorname{dom} f$.

Sous-différentiel de Fréchet $\hat{\partial}$

$$p \in \hat{\partial} f(x) \Leftrightarrow \liminf_{y \to x,\ y \neq x} \frac{f(y) - f(x) - \langle p, y - x \rangle}{\|y - x\|} \geq 0$$

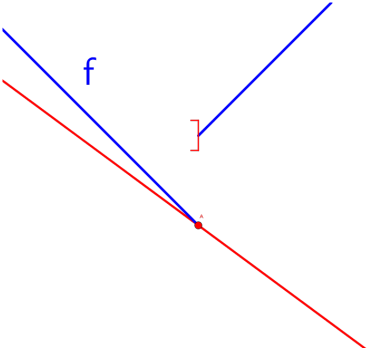


Definitions (Sous-différentiels)

$f : \mathbb{R}^N \to \mathbb{R} \cup \{+\infty\}$ s.c.i. propre. Soit $x \in \operatorname{dom} f$.

Sous-différentiel de Fréchet $\hat{\partial}$

$$p \in \hat{\partial} f(x) \Leftrightarrow \liminf_{y \to x,\ y \neq x} \frac{f(y) - f(x) - \langle p, y - x \rangle}{\|y - x\|} \geq 0$$

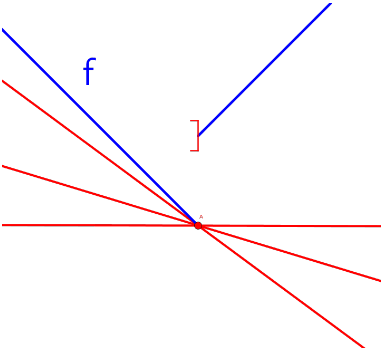


Definitions (Sous-différentiels)

$f : \mathbb{R}^N \to \mathbb{R} \cup \{+\infty\}$ s.c.i. propre. Soit $x \in \operatorname{dom} f$.

Sous-différentiel de Fréchet $\hat{\partial}$

$$p \in \hat{\partial} f(x) \Leftrightarrow \liminf_{y \to x,\ y \neq x} \frac{f(y) - f(x) - \langle p, y - x \rangle}{\|y - x\|} \geq 0$$

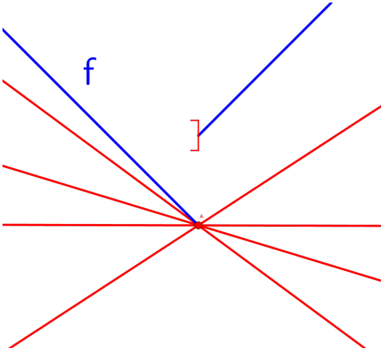


Definitions (Sous-différentiels)

$f : \mathbb{R}^N \to \mathbb{R} \cup \{+\infty\}$ s.c.i. propre. Soit $x \in \operatorname{dom} f$.

Sous-différentiel de Fréchet $\hat{\partial}$

$$p \in \hat{\partial} f(x) \Leftrightarrow \liminf_{y \to x,\ y \neq x} \frac{f(y) - f(x) - \langle p, y - x \rangle}{\|y - x\|} \geq 0$$



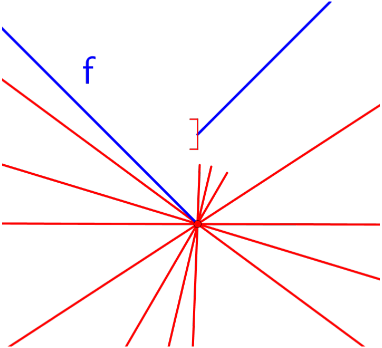
Definitions (Sous-différentiels)

$f : \mathbb{R}^N \to \mathbb{R} \cup \{+\infty\}$ s.c.i. propre. Soit $x \in \operatorname{dom} f$.

Sous-différentiel de Fréchet $\hat{\partial}$

$$p \in \hat{\partial} f(x) \Leftrightarrow \liminf_{y \to x,\ y \neq x} \frac{f(y) - f(x) - \langle p, y - x \rangle}{\|y - x\|} \geq 0$$

Ici $\hat{\partial} f(x) = [-1, +\infty[$.



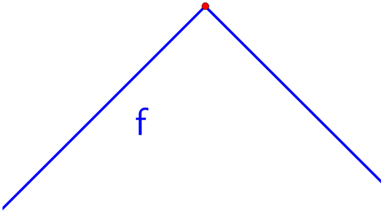
Definitions (Sous-différentiels)

$f : \mathbb{R}^N \to \mathbb{R} \cup \{+\infty\}$ s.c.i. propre. Soit $x \in \operatorname{dom} f$.

Sous-différentiel de Fréchet $\hat{\partial}$

$$p \in \hat{\partial} f(x) \Leftrightarrow \liminf_{y \to x,\ y \neq x} \frac{f(y) - f(x) - \langle p, y - x \rangle}{\|y - x\|} \geq 0$$

Ici $\hat{\partial} f(x) = \emptyset$



Definitions (Sous-différentiels)

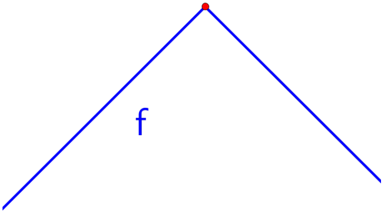
$f : \mathbb{R}^N \to \mathbb{R} \cup \{+\infty\}$ s.c.i. propre. Soit $x \in \mathrm{dom}\, f$.

Sous-différentiel Fréchet limite ∂

$$p \in \partial f(x) \quad \Leftrightarrow \quad \begin{array}{l} \exists x_n \longrightarrow x \text{ avec } f(x_n) \longrightarrow f(x), \\ \exists p_n \in \hat{\partial} f(x_n) \text{ t.q. } p_n \xrightarrow{w} p \end{array}$$

Si $x \notin \mathrm{dom}\, f$, alors $\partial f(x) = \emptyset$.

Ici $\hat{\partial} f(x) = \emptyset$



Definitions (Sous-différentiels)

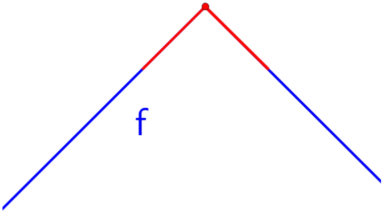
$f : \mathbb{R}^N \to \mathbb{R} \cup \{+\infty\}$ s.c.i. propre. Soit $x \in \mathrm{dom}\, f$.

Sous-différentiel Fréchet limite ∂

$$p \in \partial f(x) \quad \Leftrightarrow \quad \begin{array}{l} \exists x_n \longrightarrow x \text{ avec } f(x_n) \longrightarrow f(x), \\ \exists p_n \in \hat{\partial} f(x_n) \text{ t.q. } p_n \xrightarrow{w} p \end{array}$$

Si $x \notin \mathrm{dom}\, f$, alors $\partial f(x) = \emptyset$.

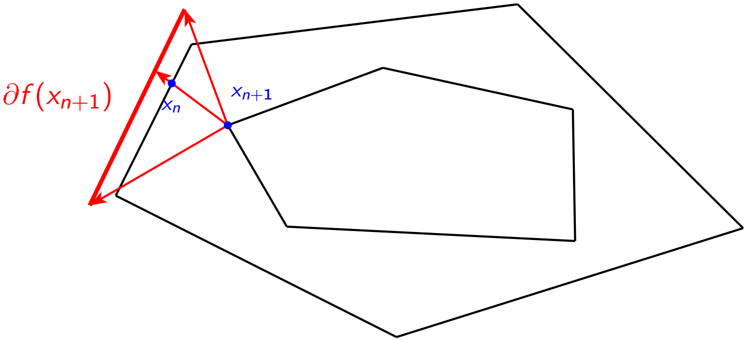
Ici $\hat{\partial} f(x) = \emptyset$ mais $\partial f(x) = \{-1, 1\}$.



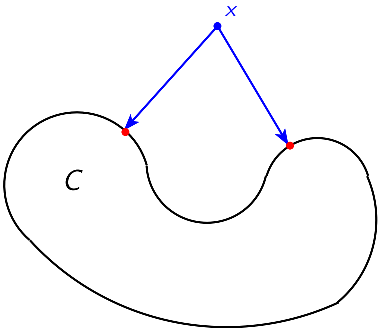
Règle de Fermat

Si x est un minimum local de f alors $0 \in \partial f(x)$ (point critique).

En particulier si $x_{n+1} \in \mathrm{prox}_f^{A_n}(x_n)$ alors $x_{n+1} \in x_n - \lambda_n \partial f(x_{n+1})$.



Soient C fermé non-vide et $f(x) = \delta_C(x) := 0$ si $x \in C, +\infty$ sinon.
Alors $prox_f^{id}(x) = \text{argmin}_{\ y \in C}\ \|y - x\|^2$
C'est la projection de x sur C.



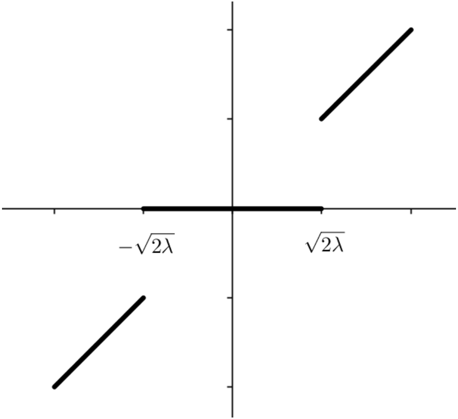
Algorithme proximal (Exemples)

Soient C fermé non-vide et $f(x) = \delta_C(x) := 0$ si $x \in C, +\infty$ sinon.
Alors $prox_f^{id}(x) = \text{argmin}_{\,y \in C} \|y - x\|^2$
C'est la projection de x sur C.

Par exemple si $H = M_{m,n}(\mathbb{R})$ et $C = \{\text{rank}\,(X) \leq r\}$
On sait que la projection se calcule en passant par une SVD et en mettant à 0 des valeurs singulières.

Algorithme proximal (Exemples)

Si $f(x) = \|x\|_0$ alors $\left(\mathrm{prox}_f^{\lambda}(x)\right)_i$:



Résultat de convergence

Soit $f : \mathbb{R}^N \longrightarrow \mathbb{R} \cup \{+\infty\}$ semi-algébrique.
Nous allons voir une classe d'algorithmes de décomposition qui vérifient :

Théorème (Frankel, G., Peypouquet, 2013 - Attouch et al., 2010)

1. Toute suite bornée est de longueur finie et convergente vers un point critique de f.
2. Si le point initial x_0 est suffisamment proche d'un minimum global a lors la suite est de longueur finie et converge vers un minimum global de f.

Soit $f = g + h : \mathbb{R}^N \longrightarrow \mathbb{R} \cup \{+\infty\}$ semi-algébrique
g s.c.i. et h de classe C^1 de gradient Lipschitzien.

$$x_{n+1} \in \mathrm{prox}_g^{A_n}\left(x_n - A_n^{-1}\nabla h(x_n)\right)$$

Forward-Backward

Soit $f = g + h : \mathbb{R}^N \longrightarrow \mathbb{R} \cup \{+\infty\}$ semi-algébrique
g s.c.i. et h de classe C^1 de gradient Lipschitzien.

$$\boxed{x_{n+1} \in \mathrm{prox}_g^{A_n}\left(x_n - A_n^{-1}\nabla h(x_n)\right)}$$

Exemple

Soit $g(x) = \delta_C(x)$ où $C \subset H$ fermé semi-algébrique et $A_n = \lambda_n^{-1} id_H$. On obtient alors l'algorithme du gradient projeté :

$$x_{n+1} \in \mathrm{proj}_C\ (x_n - \lambda_n \nabla h(x_n))$$

Soit $f = g + h : H \longrightarrow \mathbb{R} \cup \{+\infty\}$
g s.c.i. et h de classe C^1 de gradient Lipschitzien.

$$\boxed{x_{n+1} \in \mathrm{prox}_g^{A_n}\left(x_n - A_n^{-1}\nabla h(x_n)\right)}$$

Le résultat de convergence s'applique si pour $\alpha_n = \alpha(A_n)$ et $\beta_n = |||A_n|||$ on a

$$L < \underline{\alpha} \leq \alpha_n \leq \beta_n \leq \bar{\beta} < +\infty.$$

Forward-Backward

Soit $f = g + h : H \longrightarrow \mathbb{R} \cup \{+\infty\}$
g s.c.i. et h de classe C^1 de gradient Lipschitzien.

$$\boxed{x_{n+1} \in \mathrm{prox}_g^{A_n}\left(x_n - A_n^{-1}\nabla h(x_n)\right)}$$

Le résultat de convergence s'applique si pour $\alpha_n = \alpha(A_n)$ et $\beta_n = |||A_n|||$ on a

$$L < \underline{\alpha} \leq \alpha_n \leq \beta_n \leq \bar{\beta} < +\infty.$$

Dans le cas $f_{|\,\mathrm{dom}\, f}$ continue on peut prendre $\beta_n \to +\infty$ si $\frac{1}{\beta_n} \notin \ell^1$ et son taux d'accroissement reste borné.

Si f est C^2, on veut exploiter le Hessien $\nabla^2 f(x_n)$.

Si f est C^2, on veut exploiter le Hessien $\nabla^2 f(x_n)$.

Si f est strictement convexe, $A_n = \dfrac{1}{\lambda_n}\nabla^2 f(x_n)$:

(Newton) $$x_{n+1} = x_n - \lambda_n(\nabla^2 f(x_n))^{-1}\nabla f(x_n).$$

Si f est C^2, on veut exploiter le Hessien $\nabla^2 f(x_n)$.
Si f est strictement convexe, $A_n = \frac{1}{\lambda_n}\nabla^2 f(x_n)$:

(Newton) $$x_{n+1} = x_n - \lambda_n(\nabla^2 f(x_n))^{-1}\nabla f(x_n).$$

Si f est convexe, $A_n = \frac{1}{\lambda_n}(\nabla^2 f(x_n) + \varepsilon id) \quad (\varepsilon > 0)$:

(Levenberg-Marq.) $$x_{n+1} = x_n - \lambda_n(\nabla^2 f(x_n) + \varepsilon id)^{-1}\nabla f(x_n).$$

Si f n'est pas convexe, on peut prendre $A_n = \dfrac{1}{\lambda_n}(P_n + \varepsilon id)$
où P_n est une approximation positive de $\nabla^2 f(x_n)$.

Par exemple $P_n = \mathrm{proj}_{\mathcal{S}_+(\mathbb{R}^N)}(\nabla^2 f(x_n))$ en dimension finie.

Si f n'est pas convexe, on peut prendre $A_n = \dfrac{1}{\lambda_n}(P_n + \varepsilon id)$
où P_n est une approximation positive de $\nabla^2 f(x_n)$.

Par exemple $P_n = \mathrm{proj}_{\mathcal{S}_+(\mathbb{R}^N)}(\nabla^2 f(x_n))$ en dimension finie.

Et si f n'est pas C^2 mais seulement $C^{1,1}$?
Considérons la Hessienne généralisée :

$$\partial^2 f(x) = \{\lim_{n\to\infty} \nabla^2 f(x_n) \mid x_n \to x \text{ avec } f \text{ differentiable en } x_n\}.$$

On peut alors prendre n'importe quel $H_n \in \partial^2 f(x_n)$ au lieu de $\nabla^2 f(x_n)$.

La méthode se résume à :

$$H_n \in \partial^2 f(x_n)$$
$$P_n = \mathrm{proj}_{\mathcal{S}_+(\mathbb{R}^N)}(H_n)$$
$$x_{n+1} = x_n - \lambda_n (P_n + \varepsilon id)^{-1} \nabla h(x_n).$$

- $0 < \lambda_n < \bar{\lambda} < \dfrac{2\varepsilon}{L}$
- $\lambda_n \notin \ell^1$
- $\sup \dfrac{\lambda_{n+1}}{\lambda_n} < +\infty.$

Forward-Backward : Newton projeté

$f(x) = h(x) + \delta_C(x)$
avec h strictement convexe C^2 semi-algébrique
et $C \subset \mathbb{R}^N$ fermé non vide semi-algébrique. Soit $A_n = \frac{1}{\lambda_n}\nabla^2 h(x_n)$:

$$x_{n+1} \in \mathrm{proj}_C^{\nabla^2 h(x_n)}(x_n - \lambda_n(\nabla^2 h(x_n))^{-1}\nabla h(x_n)).$$

Forward-Backward : Newton projeté

$f(x) = h(x) + \delta_C(x)$
avec h strictement convexe C^2 semi-algébrique
et $C \subset \mathbb{R}^N$ fermé non vide semi-algébrique. Soit $A_n = \dfrac{1}{\lambda_n}\nabla^2 h(x_n)$:

$$x_{n+1} \in \operatorname{proj}_C^{\nabla^2 h(x_n)}(x_n - \lambda_n(\nabla^2 h(x_n))^{-1}\nabla h(x_n)).$$

$\longrightarrow$ Se généralise aussi au cas non-convexe non-C^2.

Méthode de type Gauss-Seidel

$f(x,y) = g_1(x) + g_2(y) + h(x,y)$
où g_1, g_2 s.c.i et h $C^{1,1}$ semi-algébriques.

Méthode de type Gauss-Seidel

$f(x, y) = g_1(x) + g_2(y) + h(x, y)$
où g_1, g_2 s.c.i et h $C^{1,1}$ semi-algébriques.

$$(\mathrm{GSR}) \quad \begin{cases} x_{n+1} \in \mathrm{prox}_{f(\cdot, y_n)}^{A_n}(x_n) \\ y_{n+1} \in \mathrm{prox}_{f(x_{n+1}, \cdot)}^{B_n}(y_n) \end{cases}$$

Méthode de type Gauss-Seidel

$f(x,y) = g_1(x) + g_2(y) + h(x,y)$
où g_1, g_2 s.c.i et h $C^{1,1}$ semi-algébriques.

$$
(\mathrm{GSR}) \qquad \begin{aligned} x_{n+1} &\in \mathrm{prox}_{f(\cdot,y_n)}^{A_n}(x_n) \\ y_{n+1} &\in \mathrm{prox}_{f(x_{n+1},\cdot)}^{B_n}(y_n) \end{aligned}
$$

- Etudié en 2008 (Attouch, Bolte, Soubeyran, Redont).

Méthode de type Gauss-Seidel

$f(x,y) = g_1(x) + g_2(y) + h(x,y)$
où g_1, g_2 s.c.i et h $C^{1,1}$ semi-algébriques.

$$
\text{(GSR)} \quad \begin{cases} x_{n+1} \in \mathrm{prox}^{A_n}_{f(\cdot, y_n)}(x_n) \\ y_{n+1} \in \mathrm{prox}^{B_n}_{f(x_{n+1}, \cdot)}(y_n) \end{cases}
$$

- Etudié en 2008 (Attouch, Bolte, Soubeyran, Redont).
- Ne tient pas compte de la structure lisse/non lisse.

Forward-Backward Alterné

$f(x,y) = g_1(x) + g_2(y) + h(x,y)$
où g_1, g_2 s.c.i et h $C^{1,1}$ semi-algébriques.

$$
\text{(AFB)} \quad
\begin{aligned}
x_{n+1} &\in \text{prox}_{g_1}^{A_n}\left(x_n - A_n^{-1}\nabla_x h(x_n, y_n)\right) \\
y_{n+1} &\in \text{prox}_{g_2}^{B_n}\left(y_n - B_n^{-1}\nabla_y h(x_{n+1}, y_n)\right)
\end{aligned}
$$

Forward-Backward Alterné

$f(x,y) = g_1(x) + g_2(y) + h(x,y)$
où g_1, g_2 s.c.i et h $C^{1,1}$ semi-algébriques.

$$(\text{AFB}) \quad \begin{aligned} x_{n+1} &\in \operatorname{prox}_{g_1}^{A_n}\left(x_n - A_n^{-1}\nabla_x h(x_n, y_n)\right) \\ y_{n+1} &\in \operatorname{prox}_{g_2}^{B_n}\left(y_n - B_n^{-1}\nabla_y h(x_{n+1}, y_n)\right) \end{aligned}$$

Le résultat de convergence s'applique si avec
$\alpha_n = \min\{\alpha(A_n), \alpha(B_n)\}$ et $\beta_n = \max\{|||A_n|||, |||B_n|||\}$ on a

$$L < \underline{\alpha} \leq \alpha_n \leq \beta_n \leq \bar{\beta} < +\infty.$$

Forward-Backward Alterné

$f(x,y) = g_1(x) + g_2(y) + h(x,y)$
où g_1, g_2 s.c.i et h $C^{1,1}$ semi-algébriques.

$$
\text{(AFB)} \quad
\begin{aligned}
x_{n+1} &\in \operatorname{prox}_{g_1}^{A_n}\left(x_n - A_n^{-1}\nabla_x h(x_n, y_n)\right) \\
y_{n+1} &\in \operatorname{prox}_{g_2}^{B_n}\left(y_n - B_n^{-1}\nabla_y h(x_{n+1}, y_n)\right)
\end{aligned}
$$

Le résultat de convergence s'applique si avec
$\alpha_n = \min\{\alpha(A_n), \alpha(B_n)\}$ et $\beta_n = \max\{|||A_n|||, |||B_n|||\}$ on a

$$L < \underline{\alpha} \leq \alpha_n \leq \beta_n \leq \bar{\beta} < +\infty.$$

De même si $f_{|\operatorname{dom} f}$ est continue on peut prendre $\beta_n \to +\infty$ (+hyp).

Exemple

$$(P_2) \qquad \min_{X,Y} \ \frac{1}{2}\|A - X - Y\|_F^2 + \delta_{\mathsf{rank} \leq r}(X) + \delta_{\|\cdot\|_0 \leq s}(Y)$$

$$\begin{aligned} X_{n+1} &\in \operatorname{proj}_{\{\mathsf{rank} \leq r\}} \left((1-\lambda_n)X_n - \lambda_n(Y_n - A)\right), \\ Y_{n+1} &\in \operatorname{proj}_{\{\|\cdot\|_0 \leq r\}} \left((1-\mu_n)Y_n - \mu_n(X_{n+1} - A)\right) \end{aligned}$$

Conclusion

- Algorithmes faciles à implémenter car décomposé en étapes simples
- Certificats de convergence dans un cadre non-lisse non-convexe (sans convexification)
- Fonctionne dans un cadre plus général que semi-algébrique :

$$\exists \varphi : \mathbb{R}_+ \to \mathbb{R} \text{ tq } \quad \|\partial(\varphi \circ f)(x)\| \geq 1$$

- Vitesses de convergences quadratiques pour le Newton projeté ?

Merci de votre attention !