Modèles convexes et algorithmes d'optimisation en imagerie.

Pierre Weiss.

April 21, 2011

III.1/ Théorie de la complexité en optimisation convexe. Applications à l'imagerie.

Plan de la partie

1. Eléments d'analyse convexe dans $\mathbb{R}^n$.
2. Quelques résultats en théorie de la complexité.
3. Algorithmes efficaces optimaux dans le cas différentiable et non différentiable.

Quelques références

1. T. Rockafellar *Convex Analysis* 1970.
2. B. Polyak *Introduction to optimization*, 1987.
3. D. Bertsekas *Nonlinear Programming*, 1999.
4. Y. Nesterov *Introductory lectures on optimization*, 2003.
5. A. Juditsky, cours en ligne optimisation, (ENSIMAG).
6. Bonnans, Gilbert, Lemaréchal, Sagastizàbal, *Numerical optimization*, 2006.

Notations :

On travaille dans $\mathbb{R}^n$, $n \in \mathbb{N}$ avec :

- Un produit scalaire $\langle \cdot, \cdot \rangle$. A moins que ce ne soit spécifié, il correspond au produit scalaire usuel.
- On munit l'espace d'une norme :

$$\|x\| = \sqrt{\langle x, x \rangle}.$$

Elements d'analyse convexe

Dans tout l'exposé, nous nous intéressons à des fonctions :

$$f : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}.$$

Définition (domaine d'une fonction) :

$$\mathrm{dom}(f) := \{x \in \mathbb{R}^n, f(x) < +\infty\}.$$

On suppose systématiquement que $\mathrm{dom}(f) \neq \emptyset$.
Intérêt (l'un d'eux) :

$$\inf_{x \in \mathbb{R}^n} f(x) = \inf_{x \in \mathrm{dom}(f)} f(x)$$

Les problèmes contraints s'écrivent indifféremment des problèmes non contraints.

Elements d'analyse convexe

Définition (fonction convexe) :
f est dite convexe si :

1. $\mathrm{dom}(f)$ est convexe.
2. f est convexe sur $\mathrm{dom}(f)$, $\forall (x_1, x_2) \in \mathbb{R}^n \times \mathbb{R}^n, \forall \alpha \in [0, 1]$,

$$f(\alpha x_1 + (1-\alpha)x_2) \leq \alpha f(x_1) + (1-\alpha) f(x_2).$$

Théorème :
Toute fonction f convexe en dimension finie est continue sur $\mathrm{int}(\mathrm{dom}(f))$.

Définition (épigraphe) :

$$\mathrm{epi}(f) = \{(x,t) \in \mathrm{dom}(f) \times \mathbb{R}, t \geq f(x)\}.$$

Théorème :

f est convexe si et seulement si son épigraphe est convexe.
(Schéma ?)

Un problème :
Bien qu'une fonction convexe soit continue sur $\mathrm{int}(\mathrm{dom}(f))$, elle peut avoir un comportement complexe sur le bord.
Exemple :

$$f(x,y) = \begin{cases} 0 & \text{si } x^2+y^2<1 \\ \phi(x,y) \geq 0 & \text{si } x^2+y^2=1 \\ +\infty & \text{si } x^2+y^2>1 \end{cases}$$

Minimisation : On suppose g continue sur $\mathbb{R}^2$ et on veut trouver :

$$\min_{(x,y)\in\mathbb{R}^2} f(x,y)+g(x,y)$$

Si g n'atteint pas son minimum sur le disque, il n'y a pas d'autre choix que d'explorer exhaustivement l'ensemble $\{(x,y)\in\mathbb{R}^2, x^2+y^2=1\}$!

Définition (fonction convexe fermée) :
Une fonction convexe est dite *fermée* ou *semi-continue inférieurement (s.c.i.)* si $\mathrm{epi}(f)$ est fermé.

Exemples 1D :
Voir tableau...

Définition (sous-différentiel) : Soit $f : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ convexe s.c.i. Le *sous-différentiel* de f en $x \in \mathrm{dom}(f)$ est défini par :

$$\partial f(x) = \{\eta \in \mathbb{R}^n, \forall y \in \mathbb{R}^n, f(y) \geq f(x) + \langle \eta, y - x \rangle\}.$$

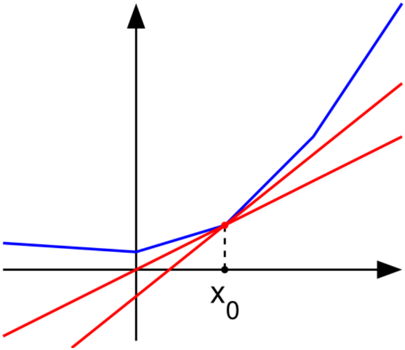


Figure: Sous-différentiel.

Propriété 1 : Si f est différentiable en x_0, alors $\partial f(x_0) = \{\nabla f(x_0)\}$.
Propriété 2 : Le sous-différentiel est non vide, convexe, fermé sur $\mathrm{int}(\mathrm{dom}(f))$.
Mais il peut être vide sur le bord du domaine.
Exemple :

$$f(x) = -\sqrt{x}$$

est convexe s.c.i. sur $[0, +\infty[$.
Et on a $\lim_{x \to 0^+} f'(x) = -\infty$ et $-\infty \notin \mathbb{R}$!

Elements d'analyse convexe

On considère le problème :

$$\text{Trouver} \quad x^* \in \operatorname*{Arg\,min}_{x \in \mathbb{R}^n} f(x)$$

où f est convexe s.c.i.

Théorème (fondamental) :

x^* est minimiseur de f si et seulement si $0 \in \partial f(x^*)$.

Une règle de calcul utile :

Soit $g(x) = f(Ax)$ où A est un opérateur linéaire, alors :

$$\partial g(x) = A^* \partial f(Ax)$$

où A^* est l'opérateur adjoint de A.

Algorithme : descente de sous-gradient (Polyak ∼ 1980)

Problème :

$$\text{Trouver} \quad x^* \in \operatorname*{Arg\,min}_{x \in X} f(x)$$

- $f : X \to \mathbb{R}$ est convexe, s.c.i.
- X est un ensemble convexe fermé.

Algorithme :

1. Choisir $x^0 \in X$ et une suite réelle de pas $(h_k)_{k \in \mathbb{N}}$ telle que :
    - $h_k \geq 0, \; \forall k \in \mathbb{N}$.
    - $\lim_{k \to +\infty} h_k = 0$.
    - $\sum_{k=0}^{+\infty} h_k = +\infty$.
2. $$x_{k+1} = \Pi_X \left(x_k - h_k \frac{\eta(x_k)}{\|\eta(x_k)\|} \right), \quad \eta(x_k) \in \partial f(x_k)$$

Algorithme : descente de sous-gradient (Polyak $\sim$ 1980)

Résultat :
Si f est L-Lipschitz sur $B = \{x \in \mathbb{R}^n, \|x_0 - x^*\| \leq R\}$, alors :

$$f_k^* - f^* \leq L \frac{R^2 + \sum_{i=0}^k h_k^2}{2 \sum_{i=0}^k h_k}$$

En particulier, si $h_k = \frac{R}{\sqrt{k+1}}$ (optimal) :

$$f_k - f^* \leq \frac{LR}{\sqrt{k+1}}$$

avec : $f_k^* = \min(f(x_0), f(x_1), ...f(x_k))$.
Ce taux est ajusté.

Algorithme : descente de sous-gradient

Note importante :
En supposant $L = R = 1$, et si on souhaite :

$$f_k^* - f^* \leq 10^{-3}$$

Dans un scénario au pire des cas, il faut 10^6 itérations !
Conclusion :

- Les descentes de sous-gradient avec pas décroissant ne doivent pas être utilisées en général.
- Elles peuvent présenter un intérêt pour coder rapidement des approximations grossières des solutions avec des pas précalculés (10-20 itérations).
- Note : les constantes L et R peuvent dépendre implicitement de la dimension n !

Optimisation convexe non différentiable : complexité

Soit $\mathcal{M}$ l'ensemble des méthodes de sous-gradient de la forme :

$$x_{k+1} \in x_0 + vect(\eta(x_0), ..., \eta(x_k)), \quad \text{où } \eta(x_i) \in \partial f(x_i).$$

Théorème : Pour tout $k \leq n-1$, pour toute méthode $m \in \mathcal{M}$, il existe une fonction $f : \mathbb{R}^n \to \mathbb{R}$

- convexe fermée.
- L-Lipschitz sur une boule de rayon R autour d'un minimiseur x^*.

telle que :

$$f(x_k) - f^* \geq \frac{MR}{1 + \sqrt{k+1}}$$

Corollaire : Les méthodes de sous-gradient avec pas en $O\left(\frac{1}{\sqrt{k}}\right)$ sont optimales.

Optimisation convexe non différentiable : complexité

Idée :
La fonction $f_k : x \mapsto \max_{1 \le i \le k} x(k) + \frac{1}{2}\|x\|^2$ est difficile à minimiser pour toutes les méthodes de sous-gradient.
Eléments de preuve :
La descente de sous-gradient avec $x_0 = 0$, assure que :

$$x_k \in \mathbb{R}^{k,n} := \{x \in \mathbb{R}^n, x(i) = 0, \forall i > k\}$$

(on ajoute qu'une coordonnée à chaque itération).
On a de plus :

$$\min_{x \in \mathbb{R}^{k,n}} f(x) - f^* \ge O\left(\frac{1}{\sqrt{k}}\right).$$

Optimisation convexe différentiable : complexité

La classe des fonctions convexes non différentiables est trop vaste pour espérer avoir des schémas génériques efficaces.

Complexité de la classe des fonctions convexes différentiables.

Soit $\mathcal{C}$, la classe des fonctions f telles que :

- f est convexe, différentiable.
- ∇f est L-Lipschitz (régularité indispensable).

Classe de méthodes :

On considère les méthodes qui génèrent des itérées du type

$$x_{k+1} \in x_0 + vect(\nabla f(x_0), ..., \nabla f(x_k)).$$

Exemples : descentes de gradient, gradient conjugué.

Optimisation convexe différentiable : complexité

Théorème : Pour tout $k \leq \frac{n-1}{2}$ et $x_0 \in \mathbb{R}^n$ il existe une fonction $f \in \mathcal{C}$ telle que :

$$f(x_k) - f^* \geq \frac{3L\|x_0 - x^*\|^2}{32(k+1)^2}$$

et

$$\|x_k - x^*\|^2 \geq \frac{1}{8}\|x_0 - x^*\|^2.$$

Conséquence :

- En général, on ne peut rien dire sur la distance au minimiseur !
- Les taux de convergence linéaires (vus en cours) $(\|x_k - x^*\| \leq \alpha^k \|x_0 - x^*\|,\ \alpha < 1)$ sont hors de portée en général.
- Le taux en $O\left(\frac{1}{k^2}\right)$ n'est pas si décourageant.

Optimisation convexe différentiable : complexité

Eléments de preuve : on exhibe la fonction la pire au monde :

$$f_k(x) = \frac{L}{4}\left(\langle A_k x, x\rangle - x(1)\right)$$

où :

$$A_k = \left[\begin{array}{cccccc} 2 & -1 & 0 & 0 & \dots & 0_{n-k,1} \\ -1 & 2 & -1 & 0 & \dots & 0_{n-k,1} \\ 0 & -1 & 2 & -1 & \dots & 0_{n-k,1} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \dots & 2 & 0_{n-k,1} \\ & & 0_{k,n-k} & & & 0_{n-k,n-k} \end{array}\right]$$

Problème : Discrétisation d'un laplacien tronqué en 1D : ce n'est rien d'autre qu'une régularisation H^1 !

Optimisation convexe différentiable : complexité

Eléments de preuve :

1. On pose $x_0 = 0$.
2. On remarque que $\nabla f(x_k) \in \mathbb{R}^{k,n}$.
3. D'où $x_k \in \mathbb{R}^{k,n}$.
4. Pour $x \in \mathbb{R}^{k,n}$, on montre :
    - $f(x) - f^* \geq O\left(\frac{L\|x_0 - x^*\|^2}{k^2}\right)$ et
    - $\|x - x^*\|^2 \geq \frac{1}{8}\|x_0 - x^*\|^2$.

Optimisation convexe différentiable : descente de gradient

On considère :

$$\min_{x\in\mathbb{R}^n} f(x), \quad \text{où } \nabla f \ \text{ est } L-\text{Lipschitz.}$$

Et la descente de gradient :

$$x_{k+1} = x_k - \tau\nabla f(x_k)$$

Preuve de convergence :

On a $\forall(x,y)\in\mathbb{R}^n\times\mathbb{R}^n$ (inégalité boom boom !):

$$\begin{aligned} f(y) &\leq \underbrace{f(x) + \langle\nabla f(x), y-x\rangle + \frac{L}{2}\|y-x\|^2}_{\psi(y,x)} && (1) \\ &= f(x) + \frac{L}{2}\|y - \left(x - \frac{\nabla f(x)}{L}\right)\|^2 - \frac{\|\nabla f(x)\|^2}{2L} && (2) \end{aligned}$$

Optimisation convexe différentiable : descente de gradient

En posant :

$$\begin{aligned} x_{k+1} &= \arg\min_{x \in \mathbb{R}^n} \psi(x, x_k) && (3) \\ &= x_k - \frac{\nabla f(x_k)}{L}, && (4) \end{aligned}$$

on assure que :

$$f(x_{k+1}) \leq f(x_k) - \frac{\|\nabla f(x_k)\|^2}{2L}.$$

On peut ensuite sommer ces inégalités de $k = 0$ à N :

$$\begin{aligned} f(x_N) - f(x_0) &\leq -\sum_{k=0}^{N} \frac{\|\nabla f(x_k)\|^2}{2L} && (5) \\ &\leq -\frac{N}{2L} \min_{1 \leq k \leq N} \|\nabla f(x_k)\|^2 && (6) \end{aligned}$$

Optimisation convexe différentiable : descente de gradient

On a donc :

$$\min_{1 \leq k \leq N} \|\nabla f(x_k)\|^2 \leq \frac{2L}{N} \cdot (f(x_0) - f^*)$$

En utilisant de plus la convexité de f (inégalité boom boom 2 !):

$$f(x^*) \geq f(x_k) + \langle \nabla f(x_k), x_k - x^* \rangle$$

et donc :

$$f(x_k) - f(x^*) \leq \|\nabla f(x_k)\| \cdot \|x_k - x_*\|$$

Optimisation convexe différentiable : descente de gradient

Résultat de convergence (relaxation):
Si f a un gradient L-Lipschitz, la descente de gradient assure que :

$$\min_{1\leq k\leq N} \|\nabla f(x_k)\|^2 \leq \frac{2L}{N} \cdot (f(x_0) - f^*)$$

si de plus f est convexe :

$$f(x_k) - f^* \leq O\left(\frac{L\|x_0 - x^*\|^2}{k}\right)$$

Ce taux de convergence est ajusté.
La méthode de gradient est sous-optimale !

Optimisation convexe différentiable : méthodes optimales

Méthodes optimales : proposées en 1983 par Y. Nesterov.

Idée générale.

Les méthodes d'optimisation à un pas ne permettent pas d'obtenir des taux de convergence optimaux ! Il faut aller au delà des principes de relaxation.

$\Rightarrow$

- Utiliser $vect\left(\{\nabla f(x_0), \nabla f(x_1), ..., \nabla f(x_k)\}\right)$ pour calculer x_{k+1}.
- Si f est convexe, ∇f apporte une information sur la topologie globale de f !
- Impératif informatique : ne pas stocker tous les gradients.

Optimisation convexe différentiable : méthodes optimales

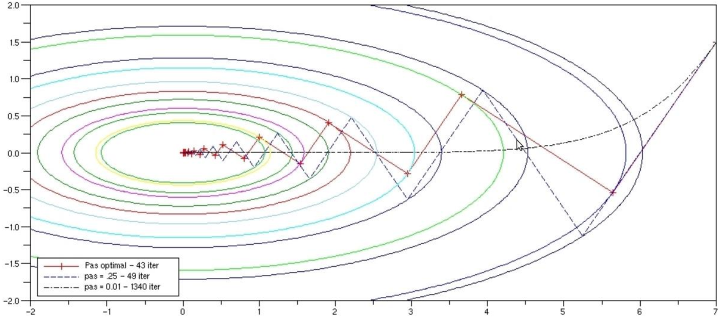


Figure: Sous-différentiel.

Optimisation convexe différentiable : méthodes optimales

Ancêtres : les méthodes "heavy-ball" (B. Polyak).

Descente de gradient oscillantes : ajouter de l'inertie !
Résoudre l'équation différentielle :

$$\ddot{x} + a\dot{x} + b\nabla f(x) = 0$$

(Solide soumis à une force de gravité et de friction)
Discrétisation :

$$\frac{-x_{k+1} + 2x_k - x_{k-1}}{\Delta t} + a\frac{(x_{k+1} - x_k)}{\Delta t} + b\nabla f(x_k) = 0$$

Soit encore :

$$x_{k+1} = x_k - a_k \nabla f(x_k) + b_k(x_k - x_{k-1})$$

(Preuve dans le cas fortement convexe à venir...)

Optimisation convexe différentiable : méthodes optimales

Idée générale

Outils :

- Une suite minimisante (x_k).
- Une suite de coefficients $A_0 = 0$, $A_{k+1} = A_k + a_k$, avec $a_k > 0$.
- Une suite de fonctions (ψ_k) approchant $A_k f$ de la forme :

$$\psi_k(x) = \sum_{i=1}^{k} a_i \left(f(x_i) + \langle \nabla f(x_i), x - x_i \rangle \right) + \frac{1}{2} \|x - x_0\|^2$$

Optimisation convexe différentiable : méthodes optimales

Maintenir les inégalités :

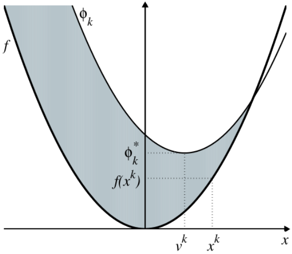
$$\begin{cases} A_k f(x_k) \leq \min_{x \in \mathbb{R}^n} \psi_k(x) \\ \psi_k(x) \leq A_k f(x) + \frac{1}{2}\|x - x_0\|^2 \quad \forall x \in \mathbb{R}^n \end{cases}$$

Ainsi, en prenant $x = x^*$, on obtient :

$$f(x_k) - f(x^*) \leq \frac{\|x^* - x_0\|^2}{A_k}$$

La rapidité de croissance de (A_k) détermine la rapidité de convergence du schéma.
Non trivial !

Optimisation convexe différentiable : méthodes optimales



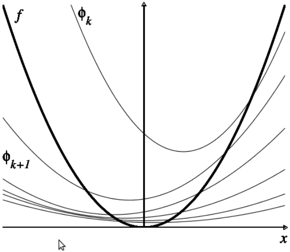


Schéma multi-pas [Nesterov 83].

- **In:** Nombre d' itérations N, point initial $x_0 \in \mathbb{R}^n$.
- **Out:** x_N une estimée de x^*.
- **Init:** Poser $t_1 = 1$, $y_1 = x_0$.

Pour k allant de 0 à N:

- Poser $x_k = y_k - \frac{\nabla f(y_k)}{L}$.
- Calculer $t_{k+1} = \frac{1+\sqrt{1+4t_k^2}}{2}$.
- Poser $y_{k+1} = x_k + \left(\frac{t_k - 1}{t_{k+1}}\right)(x_k - x_{k-1})$.

Schéma multi-pas [Nesterov 83].

Résultat de convergence

L'algorithme assure que :

$$f(x^k) - f(x^*) \leq \frac{L||x^0 - x^*||^2}{k^2}$$

C'est un taux de convergence optimal !

Distance au minimiseur.

Impossible d'obtenir des certificats sur la distance au minimiseur sous la seule hypothèse de convexité.

Définition (forte convexité).
Une fonction f est dite fortement convexe si elle est convexe et qu'il existe $\mu > 0$ tel que :

$$\forall (x, y) \in \mathrm{dom}(f)^2, \forall \eta \in \partial f(x), f(y) \geq f(x) + \langle \eta, y - x \rangle + \frac{\mu}{2} \|y - x\|^2$$

Distance au minimiseur. Une fonction fortement convexe admet un unique minimiseur x^* et

$$f(x) \geq f(x^*) + \frac{\mu}{2} \|x - x^*\|^2$$

Proposition. Une fonction C^2 est μ-fortement convexe ssi

$$\lambda_{min}(H_f(x)) \geq \mu, \forall x \in \mathrm{dom}(f).$$

Forte convexité et conditionnement.

Propriété. Soient f et g deux fonctions fortement convexes de paramètre $\mu_1 \geq 0$ et $\mu_2 \geq 0$, alors :

$$\forall(\alpha, \beta) \in \mathbb{R}_+^2, \quad x \mapsto \alpha f(x) + \beta g(x)$$

est fortement convexe de module $\alpha\mu_1 + \beta\mu_2$.

Généralisation du conditionnement.

Soit $f(x) = \frac{1}{2}\|Ax - b\|^2$

Ainsi : $\nabla f(x) = A^*(Ax - b)$.

- Constante de Lipschitz du gradient $L = \lambda_{max}(A^*A)$.
- Paramètre de forte convexité $\mu = \lambda_{min}(A^*A)$.
- Le conditionnement du système linéaire est :

$$\kappa(A^*A) = \frac{\lambda_{max}}{\lambda_{min}} = \frac{L}{\mu} = \kappa(f).$$

Forte convexité.

Quelques exemples de fonctions fortement convexes

- $f(x) = \frac{1}{2}\|x - x_0\|^2$ $(\mu = 1)$.
- $f(x) = g(x) + \frac{1}{2}\|x - x_0\|^2$ où g est convexe $(\mu = 1)$.
- $f(x) = \frac{1}{2}\|Ax - b\|^2$, $(\mu = \lambda_{min}(A^*A))$.
    - A doit être de rang plein.

Notes :

- La notion peut-être étendue à un cadre non hilbertien.
- Il peut être intéressant de changer les métriques pour faire varier ces constantes (préconditionnement).

Convergence des méthodes de gradient dans le cas fortement convexe.

Soit f une fonction μ-fortement convexe, à gradient L-Lipschitz et $\kappa = \frac{L}{\mu}$.

Théorème 1 : La descente de gradient $x_{k+1} = x_k - \frac{2}{\mu+L}\nabla f(x_k)$ assure que :

$$\|x_k - x^*\| \leq \left(\frac{\kappa - 1}{\kappa + 1}\right)^k \|x_0 - x^*\|.$$

(Arguments de point fixe).

Théorème 2 : Les schémas multi-pas avec une modification mineure du calcul des coefficients a_k assurent que :

$$\|x_k - x^*\| \leq \left(\frac{\sqrt{\kappa} - 1}{\sqrt{\kappa} + 1}\right)^k \|x_0 - x^*\|.$$

L'accélération est automatique pour les schémas simples !

Preuve de convergence.

Descente de gradient (contraction).

On a :

$$\begin{aligned}
&\|x - \tau\nabla f(x) - (y - \tau\nabla f(y))\| \\
&= \|(I - \tau\nabla f)(x - y)\| \\
&\leq \underbrace{\| \int_{t=0}^{1} (I - \tau\nabla^2 f(x + t(y - x)))(y - x)\|}_{f'(x)-f'(y)=\int_y^x f''(t)dt} \\
&\leq \underbrace{\sup_{z\in\mathbb{R}^n} |||I - \tau\nabla^2 f(z)||| \cdot \|x - y\|}_{Th.Ch.G.N.}
\end{aligned}$$

Or

$$sp(I - \tau\nabla^2 f(z)) = 1 - \tau sp(\nabla^2 f(z))$$

$$\Rightarrow \sup_{z\in\mathbb{R}^n} |||I - \tau\nabla^2 f(z)||| = \max(|1 - \tau\mu|, |1 - \tau L|)$$

Preuve de convergence.

Descente de gradient.

En prenant

$$\tau = \min_{\tau} \max(|1-\tau\mu|, |1-\tau L|) = \frac{2}{\mu + L},$$

on obtient :

$$\|(I - \tau\nabla f)(x-y)\| \leq \left(\frac{\kappa - 1}{\kappa + 1}\right) \|x - y\|$$

Ainsi :

$$\begin{aligned}
&\|x_{k+1} - x^*\| \\
&= \|x_k - \tau\nabla f(x_k) - (x^* - \tau\nabla f(x^*))\| \\
&\leq \left(\frac{\kappa - 1}{\kappa + 1}\right) \|x_k - x^*\|
\end{aligned}$$

Preuve de convergence schéma accéléré.

Idée : étudier $\|x_{k+1} - x^*\|^2 + \|x_k - x^*\|^2$

$$
\begin{aligned}
& \left\| \begin{bmatrix} x_{k+1} - x^* \\ x_k - x^* \end{bmatrix} \right\| \\
& = \left\| \begin{bmatrix} x_k - \alpha_k \nabla f(x_k) + \beta_k (x_k - x_{k-1}) - x^* \\ x_k - x^* \end{bmatrix} \right\| \\
& = \left\| \begin{bmatrix} (1+\beta_k) I & -\beta_k I \\ I & 0 \end{bmatrix} \begin{bmatrix} x_k - x^* \\ x_{k-1} - x^* \end{bmatrix} - \alpha_k \begin{bmatrix} \nabla f(x_k) \\ 0 \end{bmatrix} \right\| \\
& \leq \sup_{z \in \mathbb{R}^n} \left\| \begin{bmatrix} (1+\beta_k) I - \alpha_k \nabla^2 f(z) & -\beta_k I \\ I & 0 \end{bmatrix} \right\| \cdot \left\| \begin{bmatrix} x_k - x^* \\ x_{k-1} - x^* \end{bmatrix} \right\|
\end{aligned}
$$

Puis on résout :

$$
\min_{\alpha_k, \beta_k} \sup_{z \in \mathbb{R}^n} \left\| \begin{bmatrix} (1+\beta_k) I - \alpha_k \nabla^2 f(z) & -\beta_k I \\ I & 0 \end{bmatrix} \right\| \tag{7}
$$

Preuve de convergence schéma accéléré.

En posant :

$$\beta_k = \max(|1 - \sqrt{\alpha_k l}|, |1 - \sqrt{\alpha_k L}|)^2,$$

et

$$\alpha_k = \frac{4}{(\sqrt{L} + \sqrt{l})^2}.$$

On obtient :

$$\sup_{z \in \mathbb{R}^n} \left\| \begin{bmatrix} (1+\beta_k)I - \alpha_k \nabla^2 f(z) & -\beta_k I \\ I & 0 \end{bmatrix} \right\| \leq \left(\frac{\sqrt{\kappa} - 1}{\sqrt{\kappa} + 1} \right).$$

Preuve de convergence schéma accéléré.

Conclusion cas fortement convexe :

- Si $\kappa >> 1$, alors :

$$\left(\frac{\kappa - 1}{\kappa + 1}\right) \simeq 1 - \frac{2}{\kappa}.$$

$$\left(\frac{\sqrt{\kappa} - 1}{\sqrt{\kappa} + 1}\right) \simeq 1 - \frac{2}{\sqrt{\kappa}}.$$

- Si $\kappa >> 1$, alors les taux polynomiaux à l'origine sont plus importants que les taux linéaires asymptotiques.
- Ne pas oublier de modifier les paramètres de l'algorithme !