Théorème 0.1. *Soit f une fonction réelle α-fortement convexe et L-lipschitzienne sur E. Soit $x^* := \arg\min_{x\in E} f(x)$. On a le résultat suivant :*

$$f(\hat{x}) - f(x^*) \leq \frac{2L^2}{\alpha T} \tag{1}$$

Démonstration. Pour $t = 1$ à T, introduisons le pas :

$$\eta_t := \frac{2}{\alpha t} \tag{2}$$

En utilisant l'identité classique $2\langle x\ ,\ y\rangle = \left\|x+y\right\|^2 - \|x\|^2 - \left\|y\right\|^2$, on a pour $t = 0$ à $T-1$:

$$\begin{aligned}
f(\hat{x}) - f(x^*) &\leq f(x_t) - \left(f(x_t) + u_t^T(x^* - x_t) + \frac{\alpha}{2}\|x^* - x_t\|^2 \right) \\
&= -u_t^T(x^* - x_t) - \frac{\alpha}{2}\|x^* - x_t\|^2 \\
&= -\frac{\eta_{t+1}}{\eta_{t+1}} \cdot u_t^T(x^* - x_t) - \frac{\alpha}{2}\|x^* - x_t\|^2 \\
&= \frac{1}{2\eta_{t+1}}\left(\left\|\eta_{t+1} \cdot u_t\right\|^2 + \|x^* - x_t\|^2 - \left\|\eta_{t+1} \cdot u_t + x^* - x_t\right\|^2 \right) - \frac{\alpha}{2}\|x^* - x_t\|^2 \\
&= \frac{1}{2\eta_{t+1}}\left(\left\|\eta_{t+1} \cdot u_t\right\|^2 + \|x^* - x_t\|^2 - \left\|x^* - y_t\right\|^2 \right) - \frac{\alpha}{2}\|x^* - x_t\|^2 \\
&\leq \frac{1}{2\eta_{t+1}}(\eta_{t+1}^2 \cdot L^2 + \|x^* - x_t\|^2 - \|x^* - x_{t+1}\|^2) - \frac{\alpha}{2}\|x^* - x_t\|^2 \\
&= \frac{\eta_{t+1}L^2}{2} + \underbrace{\left(\frac{1}{2\eta_{t+1}} - \frac{\alpha}{2} \right)}_{\frac{\alpha(t-1)}{4}}\|x^* - x_t\|^2 - \underbrace{\frac{1}{2\eta_{t+1}}}_{\frac{\alpha(t+1)}{4}}\|x^* - x_{t+1}\|^2
\end{aligned}$$

En utilisant la définition de $\hat{x}$, la convexité de f et l'inégalité d'au-dessous pour tout t, on a successivement :

$$\begin{aligned}
f(\hat{x}) - f(x^*) &= f\left(\sum_{t=0}^{T} \frac{2t}{T(T+1)} x_t\right) - f(x^*) \\
&\leq \sum_{t=0}^{T} \frac{2t}{T(T+1)} \left(f(x_t) - f(x^*)\right) \\
&\leq \sum_{t=0}^{T} \frac{2t}{T(T+1)} \left(\frac{\eta_{t+1} L^2}{2} + \frac{\alpha(t-1)}{4} \cdot \|x^* - x_t\|^2 - \frac{\alpha(t+1)}{4} \cdot \|x^* - x_{t+1}\|^2\right) \\
&\leq \underbrace{\sum_{t=0}^{T} \frac{t}{T(T+1)} \eta_{t+1} L^2}_{:=A} + \underbrace{\sum_{t=0}^{T} \frac{\alpha t}{2T(T+1)} \left((t-1) \cdot \|x^* - x_t\|^2 - (t+1) \cdot \|x^* - x_{t+1}\|^2\right)}_{:=B}
\end{aligned}$$

En posant $\delta_t := t(t-1) \cdot \|x^* - x_t\|^2$, on peut réécrire la somme de gauche :

$$B = \frac{\alpha}{2T(T+1)} \sum_{t=0}^{T} (\delta_t - \delta_{t+1}) = \frac{\alpha}{2T(T+1)} (\delta_1 - \delta_{T+1}) \leq 0$$

Montrons maintenant que $A \leq \frac{2L^2}{\alpha T}$. En utilisant la définition du pas η_t, on a :

$$A = \frac{L^2}{T(T+1)} \sum_{t=0}^{T} \frac{2t}{\alpha(t+1)} = \frac{2L^2}{\alpha T} \cdot \frac{1}{(T+1)} \sum_{t=0}^{T} \underbrace{\frac{t}{t+1}}_{\leq 1} \leq \frac{2L^2}{\alpha T}$$

d'où le résultat. □

Un exemple très important est celui des fonctions α-fortement convexe et β régulière. Remarquons que nécessairement $\alpha \leq \beta$.

1 Cas de fonctions α-fortement convexe et β régulière

Soit $f : \mathbb{R}^d \mapsto \mathbb{R}$ une fonction α-fortement convexe et β régulière. On a donc pour tout x, x_0 appartenant à $\mathbb{R}^d$:

β-régularité $f(x) \leq f(x_0) + \nabla f(x_0)^T (x - x_0) + \frac{\beta}{2} \|x - x_0\|^2$

α-fortement convexe $f(x_0) + \nabla f(x_0)^T (x - x_0) + \dfrac{\alpha}{2} \|x - x_0\|^2 \leq f(x)$

Plusieurs cas :

- *Si $\alpha = \beta$, cela signifie que f est quadratique ;*
- *si $\alpha \neq \beta$, alors nécessairement $\alpha \leq \beta$.*
- *Si $\alpha \approx \beta$, alors on peut confondre la courbe de f et la courbe de β et on peut alors trouver le minimum très facilement.*

Théorème 1.1. *En reprenant les hypothèses sur f introduites au-dessus, on a le résultat suivant*

$$\|\hat{x} - x^*\|^2 \leq \mathrm{e}^{-\frac{T}{K} \cdot \|x_0 - x^*\|} \tag{3}$$

où $K := \frac{\beta}{\alpha}$.

Remarquons tout d'abord que l'on a l'égalité suivante :

$$f(\hat{x}) - f(x^*) = \int_0^1 \nabla f\left(x^* + t(\hat{x} - x^*)\right)^T \cdot (\hat{x} - x^*)\underset{.}{t} \tag{4}$$

En introduisant la fonction auxiliaire $\phi(t) := f\left(x^* + t(\hat{x} - x^*)\right)$ définie sur $[0\,;1]$, on a :

$$\phi(1) - \phi(0) = \int_0^1 \phi'(t)\underset{.}{t} = \int_0^1 \nabla f\left(x^* + t(\hat{x} - x^*)\right)^T \cdot (\hat{x} - x^*)\underset{.}{t}$$

ce qui montre bien (1) .

En utilisant la forme intégrale pour $f(\hat{x}) - f(x^*)$ et en appliquant le théorème ci dessous, on obtient donc successivement :

$$\begin{aligned}
f(\hat{x}) - f(x^*) &= \int_0^1 \nabla f\left(x^* + t(\hat{x} - x^*)\right) - \nabla f(x^*)^T \cdot (\hat{x} - x^*)\underset{.}{t} \\
&\leq \int_0^1 \|\nabla f\left(x^* + t(\hat{x} - x^*)\right) - \nabla f(x^*)\| \cdot \|\hat{x} - x^*\|\underset{.}{t} \\
&\leq \int_0^1 \beta \|t(\hat{x} - x^*)\| \cdot \|\hat{x} - x^*\|\underset{.}{t} \\
&= \beta \|\hat{x} - x^*\|^2 \cdot \int_0^1 t\underset{.}{t} \\
&= \frac{\beta \|\hat{x} - x^*\|^2}{2} \\
&\leq \frac{\beta}{2} \mathrm{e}^{-T/K} \cdot \|x_0 - x^*\|
\end{aligned}$$

Remarque à propos de la complexité de cet alogrithme :

— *il faut que l'on ne parte pas trop loin de* x^* *à cause de la dépendance au terme* $\|x_0 - x^*\|$.
— *La complexité est en* $\log(\frac{1}{\varepsilon})$ *ce qui est très rapide. Si* $\varepsilon \sim 10^{-9}$, $\log(\frac{1}{\varepsilon}) \sim 9$ *on l'on a besoin de seulement d'une dizaine d'étages pour avoir une précision à* ε*-près.*

2 Exemple de fonction à la fois α-fortement convexe et β régulière

Il suffit de prendre n'importe quelle fonction quadratique (non nulle).
Si $A \in S_d^{++}$, $b \in \mathbb{R}^d$ *et* $f : \mathbb{R}^d \to \mathbb{R}$, $x \mapsto \frac{1}{2}x^T Ax - b^T x$. *Alors on a* f *qui est :*
— $\lambda_{\min}(A)$*-fortement convexe,*
— $\lambda_{\max}(A)$*-régulière.*
En effet, on a $f \in \mathcal{L}^2$ *et* $\forall x \in \mathbb{R}^d$, $\nabla^2 f(x) = A$, *d'où l'on a bien*

$$\forall x \in \mathbb{R}^d, \quad \lambda_{\min} I_d \preceq \nabla^2 f(x) \preceq \lambda_{\max} I_d$$

En 10 étages, on peut trouver une solution à l'équation $Ax = b$ *à* ε *près, tandis que résoudre cette équation directement donne une complexité très importante.*

En effet, f *est minimale en* $\nabla f(x) = Ax - b = 0 \Leftrightarrow x = A^{-1}b$. *En appliquant l'algorithme de descente de nablaient, on trouve* $\hat{x} \in \mathbb{R}^d$ *tel que*

$$\|\hat{x} - x^*\|^2 \leq e^{-T/K} \cdot \|x_0 - x^*\|^2$$

Si $T >> \underbrace{K}_{=\frac{\lambda_{\max}}{\lambda_{\min}}} \log\left(\frac{1}{\varepsilon}\right)$, *egsi la matrice est bien conditionné (K faible), alors on approche* x^* *très rapidement.*

Démonstration. On démontre ici le théorème énoncé plus haut.

On aimerait avoir pour tout $t = 1$ à T :

$$\|x_t - x^*\|^2 \leq \left(1 - \frac{1}{K}\right) \cdot \|x_{t-1} - x^*\|^2 \tag{5}$$

ce qui impliquerait par récurrence immédiate

$$\|x_T - x^*\|^2 \leq \underbrace{\left(1 - \frac{1}{K}\right)^T}_{e^{-T/K}} \cdot \|x_0 - x^*\|^2 \tag{6}$$

On a :

$$\begin{aligned}\|x_t - x^*\|^2 &= \left\|x_{t-1} - \frac{1}{\beta}\nabla f(x_{t-1}) - x^*\right\|^2 \\ &= \|x_{t-1} - x^*\|^2 + \underbrace{\frac{1}{\beta^2}\|f(x_{t-1})\|^2 - \frac{2}{\beta}\nabla f(x_{t-1})^T \cdot (x_{t-1} - x^*)}_{\leq -\frac{1}{K}\|x_{t-1}-x^*\|^2 ?}\end{aligned}$$

Introduisons le lemme suivant :

Lemma 2.1.

$$\forall x, y \in \mathbb{R}^d \quad f\left(x - \frac{1}{\beta}\nabla f(x)\right) - f(y) \leq \nabla f(x)^T (x - y) - \frac{1}{2\beta}\|f(x)\|^2 - \frac{\alpha}{2}\left\|x - y\right\|^2 \tag{7}$$

En supposant admis le lemme et en l'appliquant à $x = x_{t-1}$ et $y = x^*$, on a :

$$\begin{aligned} 0 &\leq \nabla f(x_{t-1})^T (x_{t-1} - x^*) - \frac{1}{2\beta}\|\nabla f(x_{t-1})\|^2 - \frac{\alpha}{2}\|x_{t-1} - x^*\|^2 \\ \implies &\frac{1}{2\beta}\|\nabla f(x_{t-1})\|^2 - \nabla f(x_{t-1})^T (x_{t-1} - x^*) \leq -\frac{\alpha}{2}\|x_{t-1} - x^*\|^2\end{aligned}$$

Il suffit alors de multiplier chaque côté par $\frac{2}{\beta}$ pour avoir l'inégalité désirée.

Il ne reste plus qu'à démontrer le lemme, qui se montrer aisément en introduisant x artificiellement :

$$f(x^*) - f(y) = \underbrace{f(x^*) - f(x)}_{:=A} + \underbrace{f(x) - f(y)}_{:=B}$$

avec

$$\begin{aligned} A &\leq \nabla f(x)^T (x^+ - x) + \frac{\beta}{2}\left\|x^+ - x\right\|^2 \\ B &\leq \nabla f(x)^T (x - y) - \frac{\alpha}{2} \cdot \left\|x - y\right\|^2 \end{aligned}$$

puis il suffit d'écrire et de remplacer x^+ par sa formule.

$\square$