Questions

Exercice 1.

Soit E un ensemble convexe et compact et $f : E \to \mathbb{R}$ une fonction convexe β-régulière, où β est un paramètre connu. On considère l'algorithme 1 ci-dessous, où $\gamma_s = \frac{2}{s}$, pour tout $s = 1$ à T. Soit D le diamètre de E.

Algorithm 1 Algorithme de Frank-Wolfe

$x_0 \in \mathbb{R}^d$.
for $t = 1$ TO T **do**
  $y_{t-1} \in \underset{y \in E}{\arg\min}\, \nabla f(x_{t-1})^\top y$
  $x_t = (1-\gamma_t)x_{t-1} + \gamma_t y_{t-1}$
end for
$\hat{x} = x_T$

Question 1.

Montrer que :

$$\forall t \in \{1 \dots T\} \quad f(x_t) - f(x_{t-1}) \le \gamma_t \nabla f(x_{t-1})^\top (y_{t-1} - x_{t-1}) + \frac{\beta}{2}\gamma_t^2 D^2 \tag{1}$$

Réponse 1.

En appliquant la propriété de β-régularité aux points x_{t-1} et x_t, puis la définition de x_t, on a successivement :

$$\begin{aligned} f(x_t) - f(x_{t-1}) &\le \langle \nabla f(x_{t-1}),\, x_t - x_{t-1} \rangle + \frac{\beta}{2}\|x_t - x_{t-1}\|^2 \\ &\le \langle \nabla f(x_{t-1}),\, \gamma_t \cdot (y_{t-1} - x_{t-1}) \rangle + \frac{\beta}{2}\left\|\gamma_t \cdot (y_{t-1} - x_{t-1})\right\|^2 \end{aligned}$$

Comme $y_{t-1} \in \underset{y \in E}{\arg\min} \nabla f(x_{t-1})^\top y$ (qui existe bien car l'on optimise une fonction linéaire donc continue sur un compact), et que E est fermé, ceci assure que $y_{t-1} \in E$.

De plus, en utilisant la convexité de E, une récurrence immédiate montre que x_{t-1} appartient aussi à E. D'où

$$\left\|y_{t-1} - x_{t-1}\right\| \le \sup_{(x,y) \in E} \left\|x - y\right\| := D,$$

ce qui permet de conclure.

Question 2.

On pose $\delta_t = f(x_t) - f(x^*)$. En utilisant la définition de y_{t-1}, en déduire que :

$$\delta_t \le (1-\gamma_t)\delta_{t-1} + \frac{\beta}{2}\gamma_t^2 D^2 \tag{2}$$

Réponse 2.

Utilisons l a définition de y_{t-1} comme suggéré. On a :

$$\begin{aligned}
\forall x \in E \quad f(x_t) - f(x_{t-1}) &\leq \gamma_t \cdot \langle \nabla f(x_{t-1}), y_{t-1} - x_{t-1} \rangle + \frac{\beta \gamma_t^2}{2} D^2 \\
&\leq \gamma_t \cdot \langle \nabla f(x_{t-1}), x - x_{t-1} \rangle + \frac{\beta \gamma_t^2}{2} D^2 \\
\Leftrightarrow \forall x \in E \qquad \delta_t - \delta_{t-1} &\leq \gamma_t \cdot \langle \nabla f(x_{t-1}), x - x_{t-1} \rangle + \frac{\beta \gamma_t^2}{2} D^2
\end{aligned}$$

Pour quelle x a-t-on : $\gamma_t \cdot \langle \nabla f(x_{t-1}), x - x_{t-1} \rangle \leq -\gamma_t \cdot (f(x_t) - f(x^*))$?

Il suffit de prendre $x = x^*$ puis d'appliquer la simple convexité de f aux points x^* et x_{t+1} pour en déduire le résultat.

Question 3.

Conclure que $f(\hat{x}) - f(x^*) \leq \dfrac{\beta D^2}{T}$.

Réponse 3.

En appliquant récursivement l'équation (2), on a :

$$\begin{aligned}
f(x_T) - f(x^*) = \delta_T &\leq (1 - \gamma_T)\delta_{T-1} + \frac{\beta}{2} D^2 \gamma_T^2 \\
&\leq (1 - \gamma_T)\left\{(1 - \gamma_{T-1})\delta_{T-2} + \frac{\beta}{2} D^2 \gamma_{T-1}^2\right\} + \frac{\beta}{2} D^2 \gamma_T^2 \\
&\leq \dots \\
&\leq \delta_0 \cdot \underbrace{\prod_{k=0}^{T-1} (1 - \gamma_{T-k})}_{:=A} + \frac{\beta}{2} D^2 \underbrace{\sum_{k=1}^{T} \gamma_k^2 \underbrace{\prod_{k<s\leq T} (1 - \gamma_s)}_{:=B_k}}_{:=C}
\end{aligned}$$

Calculons les expressions séparément :

$$\begin{aligned}
A &= \prod_{k=0}^{T-1} \left(1 - \frac{2}{T-k}\right) = \frac{\prod_{k=0}^{T-1}(T-k-2)}{\prod_{k=0}^{T-1}(T-k)} = 0 \\
B_k &= \prod_{k<s\leq T} \left(1 - \frac{2}{s}\right) = \frac{\prod_{s=k+1}^{T}(s-2)}{\prod_{s=k+1}^{T} s} = \frac{(T-2)\cdot k}{(k-2)\cdot T} = \frac{k(k-1)}{T(T-1)} \\
C &= \sum_{k=1}^{T} \frac{4}{k^2} \cdot \frac{k(k-1)}{T(T-1)} \leq \sum_{k=1}^{T} \frac{4}{T(T-1)} \cdot \frac{k-1}{k} \leq \frac{4}{T}
\end{aligned}$$

D'où

$$f(\hat{x}) - f(x^*) \leq \frac{2\beta D^2}{T}$$

Question : peut-on faire mieux ?

Prenons pour s'en assurer $T = 1$ itération. On a alors $\hat{x} = x_1$, puis par l'équation (2), on a

$$\begin{aligned} f(x_1) - f(x^*) = \delta_1 &\leq (1 - \delta_1)\delta_0 + \frac{\beta}{2} \cdot 4D^2 \\ &\leq \left(1 - \frac{2}{1}\right) \cdot (f(x_0) - f(x^*)) + 2\beta D^2 \\ &\leq -\underbrace{(f(x_0) - f(x^*))}_{\geq 0} + 2\beta D^2 \leq 2\beta D^2 \end{aligned}$$

Le seul moyen d'améliorer cette inégalité serait de montrer que

$$-(f(x_0) - f(x^*)) \leq -\beta D^2,$$

ce qui me semble peu évident...