Travaux dirigés d'
Optimisation
Master 2 — Mathématiques pour les sciences du vivant

Joon Kwon

jeudi 6 octobre 2022

Exercice 1 (*Gradient stochastique*). —

1) Soit $t \geqslant 1$,

$$\left\|x^{(t+1)} - x^*\right\| = \left\|x^{(t)} - \gamma^{(t)}\hat{g}^{(t)} - x^*\right\|^2 = \left\|x^{(t)} - x^*\right\|^2 - 2\gamma^{(t)}\hat{g}^{(t)\top}(x^{(t)} - x^*) + (\gamma^{(t)})^2\left\|\hat{g}^{(t)}\right\|^2.$$

On obtient le résultat en réarrangeant les termes.

2) En sommant l'inégalité de la question 1), on obtient :

$$\sum_{t=1}^{T} \gamma^{(t)}\hat{g}^{(t)\top}(x^{(t)} - x^*) \leqslant \frac{1}{2}\left\|x^{(1)} - x^*\right\|^2 + \sum_{t=1}^{T}\frac{(\gamma^{(t)})^2}{2}\left\|\hat{g}^{(t)}\right\|^2.$$

On prend l'espérance du membre de droite et on y majore la somme de la façon suivante :

$$\mathbb{E}\left[\sum_{t=1}^{T}\frac{(\gamma^{(t)})^2}{2}\left\|\hat{g}^{(t)}\right\|^2\right] = \mathbb{E}\left[\sum_{t=1}^{T}\frac{(\gamma^{(t)})^2}{2}\mathbb{E}\left[\left\|\hat{g}^{(t)}\right\|^2 \mid x^{(t)}\right]\right] \leqslant \frac{\sigma^2}{2}\sum_{t=1}^{T}(\gamma^{(t)})^2.$$

Quant au membre gauche :

$$\begin{aligned}
\mathbb{E}\left[\sum_{t=1}^{T}\gamma^{(t)}\hat{g}^{(t)\top}(x^{(t)} - x^*)\right] &= \mathbb{E}\left[\sum_{t=1}^{T}\gamma^{(t)}\mathbb{E}\left[\hat{g}^{(t)\top}(x^{(t)} - x^*) \mid x^{(t)}\right]\right] \\
&= \mathbb{E}\left[\sum_{t=1}^{T}\gamma^{(t)}\nabla f(x^{(t)})^\top(x^{(t)} - x^*)\right] \\
&\geqslant \mathbb{E}\left[\sum_{t=1}^{T}\gamma^{(t)}(f(x^{(t)}) - f(x^*))\right].
\end{aligned}$$

où on a utilisé la convexé pour obtenir l'inégalité. On conclut en divisant par $\sum_{t=1}^{T} \gamma^{(t)}$. Si on choisit $\gamma^{(t)} = \gamma^{(1)} t^{-1/2}$ (pour $t \geqslant 1$), alors en utilisant les majoration et minoration classiques suivantes :

$$\sum_{t=1}^{T} \frac{1}{t} = 1 + \int_{1}^{T} \frac{1}{u} \,\mathrm{d}u = 1 + \log T.$$

$$\sum_{t=1}^{T} t^{-1/2} \geqslant \int_{1}^{T+1} u^{-1/2} \,\mathrm{d}u = 2(\sqrt{T+1} - 1) \geqslant \sqrt{T+1}.$$

Si σ est connu à l'avance, on peut choisir $\gamma^{(1)} = \sigma$ et il en découle la borne

$$\sigma \frac{\left\|x^{(1)} - x^*\right\|^2 + 1 + \log T}{2\sqrt{T+1}}.$$

Si σ n'est pas connu à l'avance, le choix $\gamma^{(1)} = 1$ donne la borne

$$\frac{\left\|x^{(1)} - x^*\right\|^2 + \sigma^2(1 + \log T)}{2\sqrt{T+1}}.$$

3) En utilisant la caractérisation d'ordre 1 de la régularité, on a pour tout $t \geqslant 1$,

$$f(x^{(t+1)}) - f(x^{(t)}) \leqslant \nabla f(x^{(t)})^\top (x^{(t+1)} - x^{(t)}) + \frac{L}{2} \left\|x^{(t+1)} - x^{(t)}\right\|^2 \leqslant -\gamma^{(t)} \nabla f(x^{(t)})^\top \hat{g}^{(t)} + \frac{L(\gamma^{(t)})^2}{2} \left\|\hat{g}^{(t)}\right\|^2,$$

où on a utilisé la définition du gradient stochastique pour remplacer $x^{(t+1)} - x^{(t)}$. En sommant, on obtient, en réarrangeant les termes :

$$\sum_{t=1}^{T} \gamma^{(t)} \nabla f(x^{(t)})^\top \hat{g}^{(t)} \leqslant f(x^{(1)}) - f(x^{(T+1)}) + \frac{L}{2} \sum_{t=1}^{T} (\gamma^{(t)})^2 \left\|\hat{g}^{(t)}\right\|^2.$$

En prenant l'espérance et en faisant apparaître l'espérance conditionnelle sachant $x^{(t)}$, on obtient :

$$\mathbb{E}\left[\sum_{t=1}^{T} \gamma^{(t)} \nabla f(x^{(t)})^\top \mathbb{E}\left[\hat{g}^{(t)} \mid x^{(t)}\right]\right] \leqslant f(x^{(1)}) - f(x^{(T+1)}) + \frac{L}{2} \mathbb{E}\left[\sum_{t=1}^{T} (\gamma^{(t)})^2 \mathbb{E}\left[\left\|\hat{g}^{(t)}\right\|^2 \mid x^{(t)}\right]\right],$$

ce qui donne :

$$\sum_{t=1}^{T} \gamma^{(t)} \left\|\nabla f(x^{(t)})\right\|^2 \leqslant f(x^{(1)}) - f(x^*) + \frac{L\sigma^2}{2} \sum_{t=1}^{T} (\gamma^{(t)})^2.$$

D'où le résultat en divisant par $\sum_{t=1}^{T} \gamma^{(t)}$. Si on choisit $\gamma^{(t)} = \sqrt{\frac{2}{L\sigma^2 t}}$, de façon similaire à la question précédente, on obtient :

$$\sigma \sqrt{\frac{L}{2}} \frac{f(x^{(1)}) - f(x^*) + 1 + \log T}{\sqrt{T+1}}.$$

4) En utilisant la caractérisation du premier ordre de la forte convexité, on écrit :

$$0 \geqslant f(x^*) - f(x^{(1)}) \geqslant \nabla f(x^{(t)})^\top (x^* - x^{(1)}) + \frac{K}{2} \left\| x^{(1)} - x^* \right\|^2 .$$

Ce qui donne :

$$\frac{K}{2} \left\| x^{(1)} - x^* \right\|^2 \leqslant \nabla f(x^{(1)})^\top (x^* - x^{(1)}).$$

En élevant au carré et en utilisant l'inégalité de Cauchy-Schwarz, on obtient

$$\frac{K^2}{4} \left\| x^{(1)} - x^* \right\|^2 \leqslant \left\| \nabla f(x^{(1)}) \right\|^2 ,$$

or le membre de droite peut se majorer par $\mathbb{E}\left[\left\| \widehat{g}^{(1)} \right\|^2 \right]$ car :

$$\begin{aligned} \mathbb{E}\left[\left\| \widehat{g}^{(1)} \right\|^2 \right] &= \mathbb{E}\left[\left\| \nabla f(x^{(1)}) + \widehat{g}^{(1)} - \nabla f(x^{(1)}) \right\|^2 \right] \\ &= \left\| \nabla f(x^{(1)}) \right\|^2 + \mathbb{E}\left[\left\| \widehat{g}^{(1)} - \nabla f(x^{(1)}) \right\|^2 \right] + 2 \nabla f(x^{(1)})^\top \mathbb{E}\left[\widehat{g}^{(1)} - \nabla f(x^{(1)}) \right] , \end{aligned}$$

où l'avant dernier terme est positif et le dernier terme est nul.

5) La forte convexité donne entre les points $x^{(t)}$ et x^* les inégalités suivantes.

$$\begin{aligned} f(x^*) - f(x^{(t)}) &\geqslant \nabla f(x^{(t)})^\top (x^* - x^{(t)}) + \frac{K}{2} \left\| x^{(t)} - x^* \right\|^2 \\ f(x^{(t)}) - f(x^*) &\geqslant \frac{K}{2} \left\| x^{(t)} - x^* \right\|^2 . \end{aligned}$$

En les utilisant, on écrit alors :

$$\begin{aligned} \mathbb{E}\left[\left\| x^{(t+1)} - x^* \right\|^2 \right] &= \mathbb{E}\left[\left\| x^{(t)} - x^* \right\|^2 \right] + \frac{\sigma^2}{K^2 t^2} - \frac{2}{Kt} \mathbb{E}\left[\nabla f(x^{(t)})^\top (x^{(t)} - x^*) \right] \\ &\leqslant \mathbb{E}\left[\left\| x^{(t)} - x^* \right\|^2 \right] + \frac{\sigma^2}{K^2 t^2} - \frac{2}{Kt} \mathbb{E}\left[f(x^{(t)}) - f(x^*) + \frac{K}{2} \left\| x^{(t)} - x^* \right\|^2 \right] \\ &\leqslant \mathbb{E}\left[\left\| x^{(t)} - x^* \right\|^2 \right] + \frac{\sigma^2}{K^2 t^2} - \frac{2}{Kt} \mathbb{E}\left[K \left\| x^{(t)} - x^* \right\|^2 \right] \\ &\leqslant \left(1 - \frac{2}{t} \right) \mathbb{E}\left[\left\| x^{(t)} - x^* \right\|^2 \right] + \frac{\sigma^2}{K^2 t^2}. \end{aligned}$$

6) On peut maintenant démontrer que $\mathbb{E}\left[\left\| x^{(t)} - x^* \right\|^2 \right] \leqslant \frac{4\sigma^2}{K^2 t}$ pour tout $t \geqslant 1$. Cela a déjà été montré pour $t = 1$. On procède par récurrence. Pour $t \geqslant 1$, en utilisant l'hypothèse de

récurrence,

$$\begin{aligned}
\mathbb{E}\left[\left\|x^{(t+1)} - x^*\right\|^2\right] &\leqslant \left(1 - \frac{2}{t}\right)\frac{4\sigma^2}{K^2 t} + \frac{\sigma^2}{K^2 t^2} \\
&= \frac{4\sigma^2}{K^2(t+1)}\frac{4t^2 - 3t - 7}{4t^2} \\
&\leqslant \frac{4\sigma^2}{K^2(t+1)}.
\end{aligned}$$

On conclut en utilisant la caractérisation du premier ordre de la L régularité de f :

$$\mathbb{E}\left[f(x^{(t)})\right] - f(x^*) \leqslant \mathbb{E}\left[\frac{L}{2}\left\|x^{(t)} - x^*\right\|^2\right] \leqslant \frac{2L\sigma^2}{K^2 t}.$$