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

Joon Kwon

jeudi 14 octobre 2021

Exercice 1 (*Gradient stochastique*). — Soit $f : \mathbb{R}^d \to \mathbb{R}$ une fonction différentiable admettant un minimiseur $x^* \in \mathbb{R}^d$. Soit $\sigma > 0$, $x^{(1)} \in \mathbb{R}^d$, $(\gamma^{(t)})_{t \geqslant 1}$ une suite strictement positive, et on pose pour $t \geqslant 1$:

$$x^{(t+1)} = x^{(t)} - \gamma^{(t)} \hat{g}^{(t)},$$

où $\hat{g}^{(t)}$ est une variable aléatoire telle que :

$$\mathbb{E}\left[\hat{g}^{(t)} \mid x^{(t)}\right] = \nabla f(x^{(t)}) \quad \text{et} \quad \mathbb{E}\left[\left\|\hat{g}^{(t)}\right\|^2\right] \leqslant \sigma^2.$$

1) Montrer que pour tout $t \geqslant 1$,

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

2) En déduire que pour tout $t \geqslant 1$,

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

3) On suppose dans cette question que f est convexe. Montrer que pour tout $\mathrm{T} \geqslant 1$,

$$\mathbb{E}\left[f\left(\frac{\sum_{t=1}^{\mathrm{T}} \gamma^{(t)} x^{(t)}}{\sum_{t=1}^{\mathrm{T}} \gamma^{(t)}}\right)\right] - f(x^*) \leqslant \frac{\left\|x^{(1)} - x^*\right\|^2 + \sigma^2 \sum_{t=1}^{\mathrm{T}} (\gamma^{(t)})^2}{2 \sum_{t=1}^{\mathrm{T}} \gamma^{(t)}}.$$

Proposer un choix judicieux pour $\gamma^{(t)}$, et écrire la borne qui en découle. Si σ n'est pas connu à l'avance, quelle borne peut-on obtenir ?

Soit $\mathrm{L} > 0$. Dans toute la suite de l'exercice, on suppose que f est L-régulière.

4) On ne suppose plus dans cette question que f est convexe. Montrer que pour tout $\mathrm{T} \geqslant 1$,

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

Proposer un choix judicieux pour $\gamma^{(t)}$, et écrire la borne qui en découle.

Soit $\mathrm{K} > 0$. On suppose dans la suite que f est K-fortement convexe.

5) Montrer que :

$$\left\|x^{(1)} - x^*\right\|^2 \leqslant \frac{4\sigma^2}{\mathrm{K}^2}.$$

6) On suppose dans la suite que $\gamma^{(t)} = 1/(\mathrm{K}t)$ pour tout $t \geqslant 1$. Montrer que pour tout $t \geqslant 1$,

$$\mathbb{E}\left[\left\|x^{(t+1)} - x^*\right\|^2\right] \leqslant \left(1 - \frac{2}{t}\right) \mathbb{E}\left[\left\|x^{(t)} - x^*\right\|^2\right] + \frac{\sigma^2}{\mathrm{K}^2 t^2}.$$

7) En déduire que pour tout $t \geqslant 1$,

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