Tutorial 6 : Adaptive methods - Adagrad and variants

In this tutorial, we consider a simpler version of Adagrad: Adagrad-Norm. It is defined as follows:

$$\begin{aligned} v_{k+1} &= v_k + \|\nabla f(x_k)\|^2, \qquad v_0 = 0 \\ x_{k+1} &= x_k - \alpha \frac{1}{\sqrt{v_{k+1}}} \nabla f(x_k). \end{aligned} \tag{AdaGrad-Norm}$$

Exercise 1 (Warm-up)**.** Answer the following question:

1. Is (AdaGrad-Norm) a first-order method in the Nesterov sense, i.e.:

$$x_{k+1} \in x_0 + \texttt{Lin}\left\{\nabla f(x_k), \nabla f(x_{k-1}), \dots, \nabla f(x_0)\right\}.$$

2. Consider applying (AdaGrad-Norm) to minimize the absolute value function, i.e., $f(x) = |x|$. Prove that if $x_i \neq 0, \forall i = 0, \dots, k$, we have:

$$x_{k+1} = x_k - \alpha \frac{\texttt{sign}(x_k)}{\sqrt{k+1}}.$$

3. Prove that $\lim_{k\to\infty} x_k = 0$ (Warning: this question might be complicated!!!).

Exercise 2 (First properties)**.** Consider $\{x_k\}_{k\in\mathbb{N}}$ the iterates of (AdaGrad-Norm), prove that:

1. For any $T \geq 0$ and $x \in \mathbb{R}^d$, we have:

$$\sum_{k=0}^{T-1} \langle \nabla f(x_k), x_k - x \rangle \leq \left(\frac{\max_{1\leq t\leq T-1} \|x_t - x\|^2}{2\alpha} + \alpha \right) \sqrt{v_T}.$$

Hint: you might want to use the inequalities:

$$\sum_{k=0}^{T-1} \frac{\|\nabla f(x_k)\|^2}{\sqrt{v_{k+1}}} \leq \sqrt{v_T}$$

2. Assume that $\max_{1\leq t\leq T-1} \|x_t - x\|^2 \leq R^2$, optimizing α to get the following inequalities:

$$\sum_{k=0}^{T-1} \langle \nabla f(x_k), x_k - x \rangle \leq R\sqrt{2v_T}.$$

Exercise 3 (Convex optimization problem)**.** Consider $\{x_k\}_{k\in\mathbb{N}}$ the iterates of (AdaGrad-Norm) optimizing a convex, L-Lipschitz function f with at least one minimizer $x^\star$. Prove that:

1. For any $T \geq 0$ and $x \in \mathbb{R}^d$, we have:

$$\min_{k=0,\ldots,T-1} f(x_k) - f^\star \leq \left(\frac{\max_{0\leq t\leq T-1} \|x_t - x\|^2}{2\alpha} + \alpha \right) \frac{L}{\sqrt{T}}.$$

2. Assume that $\max_{1\leq t\leq T-1} \|x_t - x\|^2 \leq R$, optimizing α to get the following inequalities:

$$\min_{k=0,\ldots,T-1} f(x_k) - f^\star \leq RL\sqrt{\frac{2}{T}}.$$

Exercise 4 (Smooth convex optimization problem)**.** Consider $\{x_k\}_{k\in\mathbb{N}}$ the iterates of (AdaGrad-Norm) optimizing a convex, L-smooth function f with at least one minimizer $x^\star$. Prove that:

1. For any $T \geq 0$ and $x \in \mathbb{R}^d$, we have:

$$\min_{k=0,\ldots,T-1} f(x_k) - f^\star \leq \left(\frac{\max_{0\leq t\leq T-1} \|x_t - x\|^2}{2\alpha} + \alpha \right)^2 \frac{2L}{T}.$$

2. Assume that $\max_{1\leq t\leq T-1} \|x_t - x\|^2 \leq R$, optimizing α to get the following inequalities:

$$\min_{k=0,\ldots,T-1} f(x_k) - f^\star \leq \frac{4R^2L}{T}.$$