Adaptive Proximal Gradient Method for Convex Optimization

Yura Malitsky   Konstantin Mishchenko

Motivation

$$\min_{x\in\mathbb{R}^d} f(x)$$

▷ f is convex and differentiable

M.-M. *Adaptive gradient descent without descent*, ICML-2020

AdGD

$$x_{k+1} = x_k - \alpha_k \nabla f(x_k)$$

$$L_k = \frac{\|\nabla f(x_k) - \nabla f(x_{k-1})\|}{\|x_k - x_{k-1}\|}$$

$$\alpha_k = \min\left\{\sqrt{1+\frac{\alpha_{k-1}}{\alpha_{k-2}}}\alpha_{k-1}\,,\ \frac{1}{2L_k}\ \right\}$$

Motivation

$$\min_{x \in \mathbb{R}^d} f(x)$$

▷ f is convex and differentiable

M.-M. *Adaptive gradient descent without descent*, ICML-2020

AdGD

$$x_{k+1} = x_k - \alpha_k \nabla f(x_k)$$

$$L_k = \frac{\|\nabla f(x_k) - \nabla f(x_{k-1})\|}{\|x_k - x_{k-1}\|}$$

$$\alpha_k = \min\left\{ \sqrt{1 + \frac{\alpha_{k-1}}{\alpha_{k-2}}}\alpha_{k-1}, \frac{1}{2L_k} \right\}$$

Questions:

- Is the first term in the update of α_k necessary?

Motivation

$$\min_{x \in \mathbb{R}^d} f(x)$$

▷ f is convex and differentiable

M.-M. *Adaptive gradient descent without descent*, ICML-2020

AdGD

$$x_{k+1} = x_k - \alpha_k \nabla f(x_k)$$

$$L_k = \frac{\|\nabla f(x_k) - \nabla f(x_{k-1})\|}{\|x_k - x_{k-1}\|}$$

$$\alpha_k = \min\left\{\sqrt{1 + \frac{\alpha_{k-1}}{\alpha_{k-2}}}\alpha_{k-1}, \frac{1}{2L_k}\right\}$$

Questions:

- Is the first term in the update of α_k necessary?
- Is 2 in the update of α_k necessary?

Motivation

$$\min_{x \in \mathbb{R}^d} f(x)$$

▷ f is convex and differentiable

M.-M. *Adaptive gradient descent without descent*, ICML-2020

AdGD

$$x_{k+1} = x_k - \alpha_k \nabla f(x_k)$$
$$L_k = \frac{\|\nabla f(x_k) - \nabla f(x_{k-1})\|}{\|x_k - x_{k-1}\|}$$
$$\alpha_k = \min\left\{ \sqrt{1 + \frac{\alpha_{k-1}}{\alpha_{k-2}}}\alpha_{k-1}, \frac{1}{2L_k} \right\}$$

Questions:

- Is the first term in the update of α_k necessary?
- Is 2 in the update of α_k necessary?
- Can we extend this algorithm to $\min f(x) + g(x)$ with prox-friendly g?

Question 1

$$
\begin{aligned}
x_{k+1} &= x_k - \alpha_k \nabla f(x_k) \\
L_k &= \frac{\|\nabla f(x_k) - \nabla f(x_{k-1})\|}{\|x_k - x_{k-1}\|} \\
\alpha_k &= \min\left\{ \sqrt{1 + \frac{\alpha_{k-1}}{\alpha_{k-2}}}\alpha_{k-1}, \frac{1}{2L_k} \right\}
\end{aligned}
$$

Question 1

$$x_{k+1} = x_k - \alpha_k \nabla f(x_k)$$

$$L_k = \frac{\|\nabla f(x_k) - \nabla f(x_{k-1})\|}{\|x_k - x_{k-1}\|}$$

$$\alpha_k = \min\left\{\sqrt{1 + \frac{\alpha_{k-1}}{\alpha_{k-2}}}\alpha_{k-1}, \frac{1}{2L_k}\right\}$$

Question 1

$$x_{k+1} = x_k - \alpha_k \nabla f(x_k)$$

$$L_k = \frac{\|\nabla f(x_k) - \nabla f(x_{k-1})\|}{\|x_k - x_{k-1}\|}$$

$$\alpha_k = \min\left\{\sqrt{1 + \frac{\alpha_{k-1}}{\alpha_{k-2}}}\alpha_{k-1}, \frac{1}{2L_k}\right\}$$

$$\alpha_k = \frac{1}{cL_k}$$

Question 1

$$x_{k+1} = x_k - \alpha_k \nabla f(x_k)$$
$$L_k = \frac{\|\nabla f(x_k) - \nabla f(x_{k-1})\|}{\|x_k - x_{k-1}\|}$$
$$\alpha_k = \min\left\{\sqrt{1 + \frac{\alpha_{k-1}}{\alpha_{k-2}}}\alpha_{k-1}, \frac{1}{2L_k}\right\}$$
$$\alpha_k = \frac{1}{cL_k}$$

Theorem. There is a convex, 1-smooth function f, such that for any $c \geqslant 1$, there is a point x_0 where this algorithm diverges.

Question 1

$$x_{k+1} = x_k - \alpha_k \nabla f(x_k)$$
$$L_k = \frac{\|\nabla f(x_k) - \nabla f(x_{k-1})\|}{\|x_k - x_{k-1}\|}$$
$$\alpha_k = \min\left\{\sqrt{1 + \frac{\alpha_{k-1}}{\alpha_{k-2}}}\alpha_{k-1}, \frac{1}{2L_k}\right\}$$
$$\alpha_k = \frac{1}{cL_k}$$

Theorem. There is a convex, 1-smooth function f, such that for any $c \geqslant 1$, there is a point x_0 where this algorithm diverges.

Answer: The first term is needed (maybe in another form).

Question 2

$$x_{k+1} = x_k - \alpha_k \nabla f(x_k)$$

$$L_k = \frac{\|\nabla f(x_k) - \nabla f(x_{k-1})\|}{\|x_k - x_{k-1}\|}$$

$$\alpha_k = \min\left\{ \sqrt{1 + \frac{\alpha_{k-1}}{\alpha_{k-2}}}\alpha_{k-1}\,,\ \frac{1}{2L_k} \right\}$$

Question 2

$$x_{k+1} = x_k - \alpha_k \nabla f(x_k)$$
$$L_k = \frac{\|\nabla f(x_k) - \nabla f(x_{k-1})\|}{\|x_k - x_{k-1}\|}$$
$$\alpha_k = \min\left\{\sqrt{1 + \frac{\alpha_{k-1}}{\alpha_{k-2}}}\alpha_{k-1}\,,\ \frac{1}{\sqrt{2}L_k}\right\}$$

- The second term $\frac{1}{2L_k}$ can be replaced by $\frac{1}{\sqrt{2}L_k}$ with exactly the same guarantees as before.

Question 2

$$
\begin{aligned}
x_{k+1} &= x_k - \alpha_k \nabla f(x_k) \\
L_k &= \frac{\|\nabla f(x_k) - \nabla f(x_{k-1})\|}{\|x_k - x_{k-1}\|} \\
\alpha_k &= \min \left\{ \sqrt{1 + \frac{\alpha_{k-1}}{\alpha_{k-2}}}\alpha_{k-1} \, , \; \frac{1}{\sqrt{2}L_k} \right\}
\end{aligned}
$$

- The second term $\frac{1}{2L_k}$ can be replaced by $\frac{1}{\sqrt{2}L_k}$ with exactly the same guarantees as before.
- The full update can be replaced by ..., which allows to use a fixed step $\alpha_k = \frac{1}{L}$.

Question 3

$$\min_{x \in \mathbb{R}^d} f(x) + g(x)$$

- f is convex and differentiable
- g is convex lsc and prox-friendly

Question 3

$$\min_{x \in \mathbb{R}^d} f(x) + g(x)$$

- f is convex and differentiable
- g is convex lsc and prox-friendly

Prox-AdGD

$$x_{k+1} = \operatorname{prox}_{\alpha_k}(x_k - \alpha_k \nabla f(x_k))$$

$$L_k = \frac{\|\nabla f(x_k) - \nabla f(x_{k-1})\|}{\|x_k - x_{k-1}\|}$$

$$\alpha_k = \min\left\{\sqrt{1 + \frac{\alpha_{k-1}}{\alpha_{k-2}}}\alpha_{k-1}, \frac{1}{\sqrt{2}L_k}\right\}$$