Constant stepsize gradient descent on smooth non-convex function

Samuel Vaiter

Created: 2023-11-09.
Last update: 2023-11-09.
Status: inprogress.

Rate for L-smooth functions

Proposition

Let $f : \mathbb{R}^d \to \mathbb{R}$ be a L-smooth function, bounded from below by $\bar{f} \in \mathbb{R}$. For any $x^{(0)}$, the GD iterates $x^{(t)}$ with $\eta^{(t)} = \eta < \frac{2}{L}$ converge towards a critical point of f,

$$\min_{0 \le t \le T} \|\nabla f(x^{(t)})\| \le \frac{1}{\sqrt{T+1}} \left(\omega^{-1} L(f(x^{(0)} - \bar{f}))\right)^{1/2},$$

where $\omega = 2\alpha(1-\alpha)$ and $\eta = \frac{2\alpha}{L}$.

We cannot say *anything* about the rate of convergence of $f(x^{(t)})$ or $(x^{(t)})_{t \ge 0}$!

Proof idea

1. Bound on one step of gradient descent.
2. Convergence of 1st-order optimality.
3. Rate of convergence.