Oracle gradient descent on coercive non-convex function

Samuel Vaiter

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

Gradient descent: oracle step-size

Gradient Descent algorithm

Require: Initialization $x^{(0)} \in \mathbb{R}^d$,
step-size policy $\eta^{(t)} > 0$.

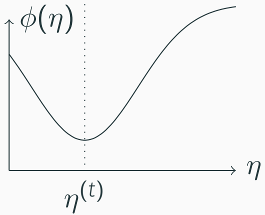
1: **for** $t = 1, \cdots$ **do**

$$x^{(t+1)} = x^{(t)} - \eta^{(t)} \nabla f(x^{(t)}) \qquad \text{(GD)}$$

- Constant step-size $\eta^{(t)} = \eta$
- Predetermined function $\eta^{(t)} = f(t)$
- "Oracle"
- Backtracking

Oracle step-size $\equiv$ best choice

$$\eta^{(t)} = \underset{\eta \geq 0}{\operatorname{argmin}} \underbrace{f(x^{(t)} - \eta \nabla f(x^{(t)}))}_{\stackrel{\text{def.}}{=} \phi(\eta)}.$$



Convergence of GD for coercive C^1 functions

Proposition

Let $f : \mathbb{R}^d \to \mathbb{R}$ be a coercive function and assume moreover that it is continuously differentiable. For any initial guess $x^{(0)}$, the oracle GD iterates $x^{(t)}$ converge up to a subsequence towards a critical point of f.

Remarks:

- *no* rate!
- If f has a unique critical point, then GD converges towards it (exercice).

Proof idea

1. The value sequence is (strictly)decreasing (unless it reaches a critical point).
2. The iterates sequence is converging (up to a subsequence).
3. The limit of the iterates sequence is a critical point.