GRADIENT DESCENT

Author
Vinitra Muralikrishnan
11th March, 2024

Contents

1 Overview

Definition 1.1. It is an algorithm for minimizing a differentiable function.

1.1 Update rule

$x_k = x_{k-1} - t_k \nabla f(x_{k-1})$. where, t_k can be fixed or adaptive.

1.2 Basic algorithm

1. choose initial point $x_0 \in \mathbb{R}$
2. repeat $x_k = x_{k-1} - t_k \nabla f(x_{k-1})$
3. stop for instnce when objective decreases by less than ϵ (user parameter).

1.3 Adaptive step size algorithm / Backtracking line search

1. Set $x_0,\ \alpha_0 > 0,\ 0 < \rho < 1$
2. On each iteration k:
3. $\quad f_k \leftarrow f(x_k)$
4. $\quad$ Set $d_k \leftarrow -\nabla f(x_{k-1})$
5. $\quad \alpha \leftarrow \alpha_0$
6. $\quad$ while $f(x_k + \alpha d_k) \geq f_k : \alpha \leftarrow \rho\alpha$
7. $\quad x_{k+1} \leftarrow x_k + \alpha d_k$

1.3.1 Exact line search

Can we choose the ideal step direction?
This would also be a minimization problem as follows:

$$t = \operatorname*{argmin}_{s \geq 0} f(x - s\nabla f(x))$$

Answer: No

- approximation is not as efficient as backtracking
- not worth solving yet another minimization problem for an existing one

2 Convergence Analysis

3 Considerations

3.1 Pros and Cons

Pros

- simple idea
- low computational cost per iterations
- fast for problems that are well-conditioned and strongly convex

Cons

- Slow for problems not strongly convex
- slow if problem not well conditioned
- only applicable to differentiable functions