

Shaper Research

February 9, 2023

1 Momentum Methods

The following is a list of accelerated methods for convergence on gradient descent. We seek to optimize objectives of the form

$$x_{t+1} = x_t - \eta \nabla f(x_t). \quad (1)$$

1.1 Polyak's Method (Heavy ball)

For some $\beta \in [0, 1]$

$$x_{t+1} = x_t - \eta \nabla f(x_t) + \beta(x_t - x_{t-1}). \quad (2)$$

1.2 Nesterov's Method (NAG)

- Polyak's has the potential to oscillate [1] despite being β -smooth and α -strongly convex.

For some $\beta \in [0, 1]$

$$x_{t+1} = x_t - \eta \nabla f(x_t + \beta(x_t - x_{t-1})) + \beta(x_t - x_{t-1}). \quad (3)$$

The primary difference between Nesterov's and Polyak's method is that Polyak's method evaluates the gradient *before* applying momentum, whereas Nesterov's method evaluates the gradient *after* applying momentum.

1.3 Adam

References

- [1] Laurent Lessard, Benjamin Recht, and Andrew Packard. Analysis and design of optimization algorithms via integral quadratic constraints. *SIAM Journal on Optimization*, 26(1):57–95, 2016.