1 The Problem

We are interested in unconstrained minimization of $f : \mathbb{R}^n \to \mathbb{R}$ given first order oracle

$$\text{minimize}_{x\in\mathbb{R}^n}\ f(x)$$

We may impose additional assumption on f, i.e. convex, L-lipschitz, μ-strongly convex

2 Gradient Descent

Gradient descent updates according to

$$x^{k+1} \leftarrow x^k - \alpha_k \nabla f(x^k)$$

for some stepsize $\alpha_k \geq 0$.

2.1 Barzilai & Borwein stepsizes

A particular choice of stepsize that relaxes the constraint on monotonic descent is given by Barzilai & Borwein [1]. The idea is to choose α_k such that $\alpha_k g^k$ approximates the Newton update.

$$\alpha_k = \frac{\langle u^k, v^k\rangle}{\|v^k\|^2} \qquad \text{or} \qquad \alpha_k = \frac{\|u^k\|^2}{\langle u^k, v^k\rangle}$$

where

$$u^k = x^k - x^{k-1} \qquad v^k = \nabla f(x^k) - \nabla f(x^{k-1})$$

3 Nesterov's Accelerated Gradient

If $f \in S^1_{L,\mu}$, then Nesterov's accelerated gradient updates according to

$$\begin{aligned} x^{k+1} &= y^k - \frac{1}{L}\nabla f(y^k) \\ y^{k+1} &= x^{k+1} + \frac{k-1}{k+2}(x^{k+1} - x^k) \end{aligned}$$

References

[1] Jonathan Barzilai and Jonathan M. Borwein. "Two-Point Step Size Gradient Methods". In: *IMA Journal of Numerical Analysis* 8.1 (Jan. 1, 1988). Publisher: Oxford Academic, pp. 141–148. ISSN: 0272-4979. DOI: 10.1093/imanum/8.1.141. URL: https://academic.oup.com/imajna/article/8/1/141/802460 (visited on 03/25/2020).