Regularized Newton Method with Global $\mathcal{O}(1/k^2)$ Convergence

Konstantin Mishchenko

جامعة الملك عبدالله للعلوم والتقنية
King Abdullah University of Science and Technology

Problem

$$\min_{x \in \mathbb{R}^d} f(x)$$

Convex and has Lipschitz Hessian

Problem

$$\min_{x \in \mathbb{R}^d} f(x)$$

$$\nabla^2 f(x) \succcurlyeq 0$$

Problem

$$\min_{x \in \mathbb{R}^d} f(x)$$

$$\nabla^2 f(x) \succcurlyeq 0$$

$$\|\nabla^2 f(x) - \nabla^2 f(y)\| \leq 2H \|x - y\|$$

Some constant

Newton and cubic Newton methods

$$x^{k+1} = x^k - (\nabla^2 f(x^k))^{-1} \nabla f(x^k)$$

Newton and cubic Newton methods

$$x^{k+1} = x^k - (\nabla^2 f(x^k))^{-1} \nabla f(x^k)$$

$$f(x) \approx f(x^k) + \langle \nabla f(x^k), x - x^k \rangle + \frac{1}{2} \langle \nabla^2 f(x^k)(x - x^k), x - x^k \rangle$$

Newton and cubic Newton methods

$$x^{k+1} = x^k - (\nabla^2 f(x^k))^{-1} \nabla f(x^k)$$

$$\begin{aligned} f(x) \leq f(x^k) &+ \langle \nabla f(x^k), x - x^k \rangle \\ &+ \frac{1}{2} \langle \nabla^2 f(x^k)(x - x^k), x - x^k \rangle \\ &+ \frac{H}{3} \|x - x^k\|^3 \end{aligned}$$

Newton and cubic Newton methods

$$x^{k+1} = x^k - (\nabla^2 f(x^k))^{-1} \nabla f(x^k)$$

$$x^{k+1} = \arg\min_x \left\{ \langle \nabla f(x^k), x - x^k \rangle + \frac{1}{2} \langle \nabla^2 f(x^k)(x - x^k), x - x^k \rangle + \frac{H}{3} \|x - x^k\|^3 \right\}$$

Newton and cubic Newton methods

$$x^{k+1} = x^k - (\nabla^2 f(x^k))^{-1} \nabla f(x^k)$$

$$x^{k+1} = \arg\min_x \left\{ \langle \nabla f(x^k), x - x^k \rangle + \frac{1}{2} \langle \nabla^2 f(x^k)(x - x^k), x - x^k \rangle + \frac{H}{3} \|x - x^k\|^3 \right\}$$

Nesterov & Polyak, 2006

Griewank, 1981

Newton and cubic Newton methods

$$x^{k+1} = x^k - (\nabla^2 f(x^k))^{-1} \nabla f(x^k)$$

$$x^{k+1} = x^k - (\nabla^2 f(x^k) + H\|x^{k+1} - x^k\|\mathbf{I})^{-1} \nabla f(x^k)$$

Newton and cubic Newton methods

$$\sout{x^{k+1} = x^k - (\nabla^2 f(x^k))^{-1} \nabla f(x^k)}$$

$$x^{k+1} = x^k - (\nabla^2 f(x^k) + H\|x^{k+1} - x^k\|\mathbf{I})^{-1} \nabla f(x^k)$$

Lemma. It holds $H\|x^{k+1} - x^k\| \approx \sqrt{H\|\nabla f(x^{k+1})\|}$.

Newton and cubic Newton methods

$$\sout{x^{k+1} = x^k - (\nabla^2 f(x^k))^{-1} \nabla f(x^k)}$$

$$x^{k+1} = x^k - (\nabla^2 f(x^k) + H\|x^{k+1} - x^k\|\mathbf{I})^{-1} \nabla f(x^k)$$

Lemma. It holds $H\|x^{k+1} - x^k\| \approx \sqrt{H\|\nabla f(x^{k+1})\|}$.

Newton and cubic Newton methods

$$\sout{x^{k+1} = x^k - (\nabla^2 f(x^k))^{-1} \nabla f(x^k)}$$

$$x^{k+1} = x^k - (\nabla^2 f(x^k) + H\|x^{k+1} - x^k\| \mathbf{I})^{-1} \nabla f(x^k)$$

Lemma. It holds $H\|x^{k+1} - x^k\| \approx \sqrt{H\|\nabla f(x^{k+1})\|}$.

Idea. Maybe $H\|x^{k+1} - x^k\| \approx \sqrt{H\|\nabla f(x^k)\|}$?

Newton and cubic Newton methods

$$\sout{x^{k+1} = x^k - (\nabla^2 f(x^k))^{-1} \nabla f(x^k)}$$

$$x^{k+1} = x^k - (\nabla^2 f(x^k) + H\|x^{k+1} - x^k\|\mathbf{I})^{-1} \nabla f(x^k)$$

Lemma. It holds $H\|x^{k+1} - x^k\| \approx \sqrt{H\|\nabla f(x^{k+1})\|}$.

Idea. Maybe $H\|x^{k+1} - x^k\| \approx \sqrt{H\|\nabla f(x^k)\|}$?

NO!

Newton and cubic Newton methods

$$\sout{x^{k+1} = x^k - (\nabla^2 f(x^k))^{-1} \nabla f(x^k)}$$

$$x^{k+1} = x^k - (\nabla^2 f(x^k) + H\|x^{k+1} - x^k\|\mathbf{I})^{-1} \nabla f(x^k)$$

Lemma. It holds $H\|x^{k+1} - x^k\| \approx \sqrt{H\|\nabla f(x^{k+1})\|}$.

Idea. Maybe $H\|x^{k+1} - x^k\| \approx \sqrt{H\|\nabla f(x^k)\|}$?

NO! **But the method works!**

Proposed method

$$x^{k+1} = x^k - \left(\nabla^2 f(x^k) + \sqrt{H\|\nabla f(x^k)\|}\mathbf{I}\right)^{-1} \nabla f(x^k)$$

Proposed method

$$x^{k+1} = x^k - \left(\nabla^2 f(x^k) + \sqrt{H\|\nabla f(x^k)\|}\mathbf{I}\right)^{-1} \nabla f(x^k)$$

Everything is known!

Proposed method

$$x^{k+1} = x^k - \left(\nabla^2 f(x^k) + \sqrt{H\|\nabla f(x^k)\|\mathbf{I}}\right)^{-1} \nabla f(x^k)$$

Everything is known!

$$f(x^k) - f(x^*) = \mathcal{O}\left(\frac{1}{k^2}\right)$$

Proposed method

$$x^{k+1} = x^k - \left(\nabla^2 f(x^k) + \sqrt{H\|\nabla f(x^k)\|}\mathbf{I}\right)^{-1} \nabla f(x^k)$$

Everything is known!

$$f(x^k) - f(x^*) = \mathcal{O}\left(\frac{1}{k^2}\right)$$

↑

Convex and has Lipschitz Hessian

Where to find the paper

Beyond First-Order Methods in ML
ICML 2021

https://sites.google.com/view/optml-icml2021