Lecture 4

Thursday, March 11, 2021 3:26 PM

$$\min_{x \in \mathbb{R}^n} f(x), \quad f : \mathbb{R}^n \to \mathbb{R}$$

- All the first order algorithms that we saw had a $\mathrm{poly}(1/\varepsilon)$ dependence (running time) on the error ε.
- Start the journey towards understanding algorithms with $\mathrm{polylog}(1/\varepsilon)$ running time dependence on ε.

Newton's method — uses second order oracle access to f.

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

$H_f(x_t)$: Hessian of f at x_t

$H_f(x)$: $n \times n$ matrix, $\quad H_f(x)_{ij} = \frac{\partial^2 f}{\partial x_i \partial x_j}$

Fact: f is convex iff $H_f(x)$ is PSD $\forall x$.

$$x_{t+1} := \arg\min_x \; f(x_t) + \langle \nabla f(x_t), x - x_t \rangle + \frac{1}{2} \langle x - x_t, H_f(x_t)(x - x_t) \rangle$$

$$x_{t+1} := x_t - H_f(x_t)^{-1} \nabla f(x_t) \qquad \text{Newton's iteration}$$

Analysis in the Euclidean norm:

Assumptions: ① Strong convexity: $H_f(x) \succcurlyeq h I \quad \forall x$

(or $\lambda_{\min}(H_f(x)) \geq h$)

② Lipschitzness of the Hessian (third order smoothness): $\| H_f(x) - H_f(y) \| \leq L \|x - y\|_2 \quad \forall x, y$

Theorem: $x_1 := x_0 - H_f(x_0)^{-1} \nabla f(x_0)$

Theorem: $\quad x_1 := x_0 - H_f(x_0) \quad \nabla f(x_0)$

$$\Rightarrow \quad \| x_1 - x^* \|_2 \leq \frac{L}{2h} \| x_0 - x^* \|_2^2$$

(quadratic convergence)

Issue: Not many functions for which can get good bounds on $L/2h$.

Example: $f(x_1, x_2) := -\log(c_1 - x_1) - \log(c_1 + x_1) - \log(c_2 - x_2) - \log(c_2 + x_2)$

For various values of c_1, c_2, the above bound does not predict the convergence of Newton's method

Local norms: $f : \mathbb{R}^n \to \mathbb{R}$ strictly convex function i.e.
$H_f(x)$ is positive definite (PD) $\quad \forall\, x \in \mathbb{R}^n$

Local inner product: $\quad \langle u, v \rangle_x = \langle u, H_f(x)\, v \rangle$

Local norm: $\quad \|u\|_x = \sqrt{\langle u, H_f(x)\, u \rangle}$

forms a Riemannian metric $\quad$ FYI

Gradient flow: steepest descent w.r.t. Euclidean norm i.e.

$$\rightarrow \underset{u:\, \|u\|_2 = 1}{\operatorname{argmin}} \; \frac{d}{dt} f(x + tu) \Big|_{t=0} = -\frac{\nabla f(x)}{\|\nabla f(x)\|_2}$$

What about steepest descent w.r.t. the local norms? i.e.

Claim: $$\underset{u:\, \|u\|_x = 1}{\operatorname{argmin}} \; \frac{d}{dt} f(x + tu) \Big|_{t=0} = -\frac{H_f(x)^{-1} \nabla f(x)}{\|H_f(x)^{-1} \nabla f(x)\|_x}$$

Pf:

$$\underset{u:\ \|u\|_x=1}{\operatorname{argmin}} \langle \nabla f(x), u \rangle$$

$$\downarrow$$

$$\underset{u:\ \langle u, H_f(x) u\rangle = 1}{\operatorname{argmin}} \langle \nabla f(x), u \rangle$$

$$\left| \langle H_f(x)^{-1/2} \nabla f(x), H_f(x)^{-1/2} u \rangle \right|$$

$$\leq \| H_f(x)^{-1/2} \nabla f(x) \|_2 \underbrace{\| H_f(x)^{-1/2} u \|_2}_{=1}$$

$$= \| H_f(x)^{-1/2} \nabla f(x) \|_2$$

For $u = \dfrac{-H_f(x)^{-1} \nabla f(x)}{\| H_f(x)^{-1} \nabla f(x) \|_x}$, $\|u\|_x = 1$ by construction

$$\langle \nabla f(x), u \rangle = -\frac{\langle \nabla f(x), H_f(x)^{-1} \nabla f(x) \rangle}{\| H_f(x)^{-1} \nabla f(x) \|_x}$$

$$= -\frac{\| H_f(x)^{-1/2} \nabla f(x) \|_2^2}{\sqrt{\langle H_f(x)^{-1} \nabla f(x), H_f(x) H_f(x)^{-1} \nabla f(x) \rangle}}$$

$$= - \| H_f(x)^{-1/2} \nabla f(x) \|_2$$

Newton's increment:

$$x_{t+1} := x_t - \underbrace{H_f(x_t)^{-1} \nabla f(x_t)}_{n(x_t)}$$

Potential function: $\|n(x)\|_x$

Assumption: Hessians at x & y are close if x & y close

Let $\delta_0 < 1$ be some fixed constant.

Definition: f satisfies the Newton-Local (NL) Condition for $\delta_0 < 1$ if $\forall\ 0 \le \delta \le \delta_0$, $\forall\ x, y$ s.t. $\|y-x\|_x \le \delta$,

$$(1-3\delta)\, H(x) \preceq H(y) \preceq (1+3\delta)\, H(x)$$

$(A \preceq B$ if $B - A \succeq 0)$

Theorem: Let $f: \mathbb{R}^n \to \mathbb{R}$ be a strictly convex function satisfying the NL Condition for $\delta_0 = 1/6$. Then

$$x_1 := x_0 - n(x_0)$$

s.t. $\|n(x_0)\|_{x_0} \le 1/6$

$$\Rightarrow \|n(x_1)\|_{x_1} \le 3\, \|n(x_0)\|_{x_0}^2$$

Lemma: Suppose x, y s.t. $\|y-x\|_x \le 1/6$. Then

the x & y local norms are close i.e. $\forall\ u$

① $\frac{1}{2}\|u\|_x \le \|u\|_y \le 2\|u\|_x$

② $\frac{1}{2}\|u\|_{H_f(x)^{-1}} \le \|u\|_{H_f(y)^{-1}} \le 2\|u\|_{H_f(x)^{-1}}$

$$\left(\|u\|_A = \sqrt{\langle u, Au \rangle} \right)$$

<u>Pf:</u> From the NL condition, we have that

$$\frac{1}{2} H_f(y) \preceq H_f(x) \preceq 2 H_f(y)$$

$$\frac{1}{2} H_f(y)^{-1} \preceq H_f(x)^{-1} \preceq 2 H_f(y)^{-1}$$

$$\left(A \preceq B \Rightarrow B^{-1} \preceq A^{-1} \right)$$

e.g.

$$\Rightarrow \forall u \ \frac{1}{2} \langle u, H_f(y) u \rangle \leq \langle u, H_f(x) u \rangle$$

$$\Rightarrow \|u\|_y \leq 2 \|u\|_x$$