Lecture 5

Tuesday, March 16, 2021 9:29 AM

Recall:

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

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

PSD ordering

Theorem [analysis of Newton's iteration]: Let $f: \mathbb{R}^n \to \mathbb{R}$ be a strictly convex function that satisfies the NL condition for $\delta_0 = 1/6$.

Then if $x_1 := x_0 + n(x_0)$

(recall $n(x_0) := -H_f(x_0)^{-1} \nabla f(x_0)$)

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

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

Already proved in last lecture:

Lemma: Suppose x, y are such that $\|y-x\|_x \le 1/6$ (Also f satisfies the NL condition with $\delta_0 = 1/6$). Then $\forall u$

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

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

Proof of Theorem [analysis of Newton's iteration]:

$$\|n(x_1)\|_{x_1} = \sqrt{\langle n(x_1) , H_f(x_1)\, n(x_1)}$$

$$= \sqrt{\left\langle H_f(x_1)^{-1} \nabla f(x_1) , H_f(x_1)\, H_f(x_1)^{-1} \nabla f(x_1) \right\rangle}$$

$$= \|\nabla f(x_1)\|_{H_f(x_1)^{-1}}$$

$$\|n(x_0)\|_{x_0} = \|\nabla f(x_0)\|_{H_f(x_0)^{-1}}$$

Because of the lemma, it suffices to prove that

$$\|\nabla f(x_1)\|_{H_f(x_0)^{-1}} \leq 3/2\, \|\nabla f(x_0)\|_{H_f(x_0)^{-1}}$$

Fundamental theorem of calculus: $g(1) = g(0) + \int_0^1 g'(t)\, dt$

$$h(t) = \nabla f(x_0 + t(x_1 - x_0))$$

$$\nabla f(x_1) = \nabla f(x_0) + \int_0^1 H_f(x_0 + t(x_1 - x_0))(x_1 - x_0)\, dt$$

$$= \nabla f(x_0) - \int_0^1 H_f(x_0 + t(x_1 - x_0))\, H_f(x_0)^{-1} \nabla f(x_0)\, dt$$

$$= \underbrace{\left(H_f(x_0) - \int_0^1 H_f(x_0 + t(x_1 - x_0))\, dt \right)}_{M(x_0)} H_f(x_0)^{-1} \nabla f(x_0)$$

$$\|\nabla f(x_1)\|_{H_f(x_0)^{-1}} = \sqrt{\left\langle M(x_0)\, H_f(x_0)^{-1} \nabla f(x_0) , H_f(x_0)^{-1} M(x_0)\, H_f(x_0)^{-1} \nabla f(x_0) \right\rangle}$$

$$= \sqrt{\langle \underbrace{H_f(x_0)^{-1/2} M(x_0) H_f(x_0)^{-1/2}}_{N(x_0)} H_f(x_0)^{-1/2} \nabla f(x_0), \underbrace{H_f(x_0)^{-1/2} M(x_0) H_f(x_0)^{-1/2}}_{} H_f(x_0)^{-1/2} \nabla f(x_0) \rangle}$$

$$= \| N(x_0) H_f(x_0)^{-1/2} \nabla f(x_0) \|_2$$

$$\leq \| N(x_0) \| \; \| H_f(x_0)^{-1/2} \nabla f(x_0) \|_2$$

$$= \| N(x_0) \| \; \| \nabla f(x_0) \|_{H_f(x_0)^{-1}}$$

To complete the proof, need $\| N(x_0) \| \leq \frac{3}{2} \| \nabla f(x_0) \|_{H_f(x_0)^{-1}}$

$$N(x_0) = H_f(x_0)^{-1/2} M(x_0) H_f(x_0)^{-1/2}$$

[Easy to prove fact: $\| A^{-1/2} B A^{-1/2} \| \leq \alpha$ iff $-\alpha A \preceq B \preceq \alpha A$]

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

suffices to prove that $-\frac{3}{2}\delta H(x_0) \preceq M(x_0) \preceq \frac{3}{2}\delta H(x_0)$

$$M(x_0) = H(x_0) - \int_0^1 H(x_0 + t(x_1 - x_0))\, dt$$

$$z_t := x_0 + t(x_1 - x_0)$$

$$\| z_t - x_0 \|_{x_0} = t \| x_1 - x_0 \|_{x_0} = t\delta$$

$$\Rightarrow (1 - 3t\delta) H(x_0) \preceq H(z_t) \preceq (1 + 3t\delta) H(x_0)$$

(NL condition)

$$\Rightarrow \quad (1-3t\delta)H(x_0) \leq H(z_t) \leq (1+3t\delta)H(x_0)$$ (NL condition)

$$\Rightarrow \quad -3t\delta\, H(x_0) \leq H(x_0) - H(z_t) \leq 3t\delta\, H(x_0)$$

Integrating on t from 0 to 1 finishes the proof.

Interior point methods for Linear Programming

$$P = \{ x \in \mathbb{R}^n : \langle a_i, x\rangle \leq b_i \quad \forall\, i \in [m] \}$$

$Ax \leq b$

Linear program: $\min \langle c, x\rangle$ s.t. $x \in P$

Theorem: There is an algorithm that given the description of the LP (A,b,c) with n variables, m constraints & total bit complexity L, $\varepsilon > 0$, & assuming that the polytope P is full-dimensional & non-empty, outputs $\hat{x} \in P$ s.t.

$$\langle c, \hat{x}\rangle \leq \langle c, x^*\rangle + 2\varepsilon$$

(x^* optimal solution)

(running time $\mathrm{poly}(L, \log(1/\varepsilon))$)

Remarks:

1. Possible to remove the full-dimensionality assumption.
2. Possible to solve the LP exactly. Set $\varepsilon = 2^{-\mathrm{poly}(L)}$ & then do some rounding.

Barrier functions: $K \subseteq \mathbb{R}^n$ convex set.

$F : \text{int}(K) \to \mathbb{R}$ is said to be a barrier function for K if F approches $+\infty$ as we approach the boundary of K. Also require F to be strictly convex.

Example: $K \subseteq \mathbb{R} = \{ x : x \geq 0 \}$

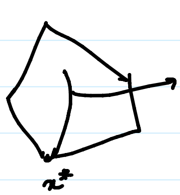
$F(x) = 1/x$

An approach for $\min \langle c, x \rangle$ s.t. $x \in K$
↓
Consider a family of objective functions

$$f_\eta(x) := \eta \langle c, x \rangle + F(x)$$

as η gets big enough, the optimal solution for f_η gets close the optimal sol^n x^* for the above program

$$x^*_\eta := \underset{x \in \text{int}(K)}{\text{argmin}} \ \eta \langle c, x \rangle + F(x)$$



$\to$ curve of x^*_η
(central path)

Path following Interior Point Methods (IPM)

$$P = \{ x \in \mathbb{R}^n : \langle a_i, x \rangle \leq b_i \ \forall i \in [m] \}$$

Definition [Logarithmic barrier] : $F(x) := -\sum_{i=1}^{m} \log\left(b_i - \langle a_i, x\rangle\right)$

(defined only in the interior of P)

The idea is to follow the central path.

1. **Initialization** : How to find some initial $\eta > 0$ & a point close to x^*_η ?
2. **Following the central path** : How to follow the central path using discrete steps?
3. **Termination** : When to stop? How to guarantee that we have reached close to x^* ?

Suppose we are given x_0 close to x_{η_0} for some $\eta_0 > 0$.
We will use Newton's method to update the current point.

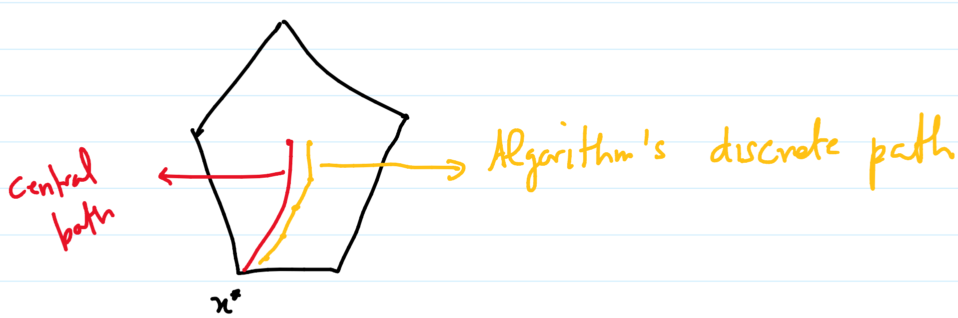
$$x_1 := x_0 + N_{\eta_0}(x_0), \qquad N_\eta(x) := -H_{f_\eta}(x)^{-1} \nabla f_\eta(x)$$

first step

$$= -H_F(x)^{-1} \nabla f_\eta(x)$$

more generally $x_{t+1} := x_t + N_{\eta_t}(x_t)$

& then also update $\eta_{t+1} := \eta_t (1+\gamma)$



What is the notion of closeness to maintain?
whatever came out of the analysis of Newton's method.

Invariant that we'll maintain: $\|N_{\eta_t}(x_t)\|_{x_t} := \|N_{\eta_t}(x_t)\|_{H_F(x_t)} \leq 1/6$

IPM Algorithm:

1. Initialization: Find some $\eta_0 > 0$ & x_0 s.t. $\|N_{\eta_0}(x_0)\|_{x_0} \leq 1/6$.
2. T s.t. $\eta_T := \left(1 + \frac{1}{20\sqrt{m}}\right)^T \eta_0 > m/\varepsilon$

for $t = 0, 1, \ldots T$

3. Newton step: $x_{t+1} := x_t + N_{\eta_t}(x_t)$
4. Update η: $\eta_{t+1} := \eta_t \left(1 + 1/20\sqrt{m}\right)$

Termination: $\hat{x} :=$ Two Newton steps from x_T w.r.t. f_{η_T}.

return $\hat{x}$

Theorem: Above algorithm after $T = O\left(\sqrt{m} \log\left(\frac{m}{\varepsilon \eta_0}\right)\right)$ iterations outputs a point $\hat{x}$ s.t.

$$\langle c, \hat{x} \rangle \leq \langle c, x^* \rangle + 2\varepsilon$$

Each iteration requires solving a linear system of the kind $H_F(x)\, y = z$ (y is the unknown).