IPM

Sunday, March 30, 2025 10:20 AM

Yulan

$$\min_{x \in K} c^T x \quad \rightarrow \quad \min \; t\, c^T x + \phi(x)$$

$\phi(x)$: convex

barrier for K

self-concordant $\left| D^3 \phi(x)[h,h,h] \right| \le 2 \left(D^2 \phi(x)[h,h] \right)^{3/2}$

$$\left(g(t) = \phi(x+th) \quad \left| g'''(t) \right| \le 2 \left(g''(t) \right)^{3/2} . \right)$$

$$\|\nabla \phi(x)\|^2_{\nabla^2 \phi(x)^{-1}} \le \nu \quad \longleftarrow \text{ barrier parameter.}$$

$\phi(x) = -\ln x$

$$-\sum \ln x_i \,, \quad -\sum_i \ln (b_i - a_i^T x) \,, \quad -\ln \det X .$$

Easy to optimize by Newton Iteration.

Lemma 1. For self-concordant f,

if $\|\nabla f(x)\|_{\nabla^2 f(x)^{-1}} = r < 1$, then $\hat{x} \leftarrow x - \nabla^2 f(x)^{-1} \nabla f(x)$

has $\|\nabla f(\hat{x})\|_{\nabla^2 f(\hat{x})^{-1}} \le \dfrac{r^2}{1-r}$.

Note: if ϕ has barrier parameter ν

then $\frac{\phi}{2}$ ——————— $\frac{\nu}{2}$

Proof uses the following crucial observation that Hessian of a self-concordant function changes slowly.

Lemma 2 $\forall x, y$ $\|x-y\|_x, \|x-y\|_y < 1$,

$$(1-\|x-y\|_x^2)\nabla^2 f(x) \preceq \nabla^2 f(y) \preceq \frac{\nabla^2 f(x)}{(1-\|x-y\|_x^2)}$$

Pf. Let $\alpha(t) = \langle \nabla^2 f(x+t(y-x))u, u \rangle$.

$$\alpha'(t) = D^3 f(x+t(y-x))[y-x, u, u]$$

By s.c. $|\alpha'(t)| \le 2\|y-x\|_{x+t(y-x)} \|u\|^2_{x+t(y-x)}$ (*)

For $u = y-x$

$$|\alpha'(t)| \le 2\,\alpha(t)^{3/2}$$

$$\frac{d}{dt}\left(\frac{1}{\sqrt{\alpha(t)}}\right) \ge -1$$

$\left(\frac{d}{dt}\frac{1}{\sqrt{x}} = -\frac{1}{2}\frac{1}{x^{3/2}}\right.$

$$\frac{1}{\quad} \ge \frac{1}{\quad} - t = \frac{1}{\|u\|} - t$$

$$\frac{1}{\sqrt{\alpha(t)}} \geq \frac{1}{\sqrt{\alpha(0)}} - t = \frac{1}{\|y-x\|_x} - t$$

$$\|y-x\|_{x+t(y-x)} \leq \frac{\|y-x\|_x}{(1 - t\|y-x\|_x)}$$

$$\therefore (*) \Rightarrow \quad |\alpha'(t)| \leq \frac{2\|y-x\|_x}{(1-t\|y-x\|_x)} \alpha(t)$$

$$\left|\frac{d}{dt} \ln \alpha(t)\right| \leq \frac{2\|y-x\|_x}{(1-t\|y-x\|_x)} = \frac{d}{dt}\left(-2\ln(1-t\|y-x\|_x)\right)$$

Integrating from $t=0$ to 1:

$$(1-\|y-x\|_x)^2 \leq \frac{\alpha(t)}{\alpha(0)} \leq (1-\|y-x\|_x)^{-2}$$

Pf. (L1). By L2, $\|\nabla f(\hat{x})\|_{\nabla^2 f(\hat{x})^{-1}} \leq \frac{\|\nabla f(\hat{x})\|_{\nabla^2 f(x)^{-1}}}{(1-r)}$

(Since $\|\hat{x}-x\|_x^2 = (\nabla^2 f(x)^{-1}\nabla f(x))^T \nabla^2 f(x) \nabla^2 f(x)^{-1} \nabla f(x)$

$$= \|\nabla f(x)\|^2_{\nabla^2 f(x)^{-1}} = r^2$$

we have $\nabla^2 f(\hat{x})^{-1} \preceq (1-r)^{2} \nabla^2 f(x)^{-1}$).

$$\nabla f(\hat{x}) = \nabla f(x) + \int_0^1 \nabla^2 f(x+t(\hat{x}-x))(\hat{x}-x)\,dt$$

$$= \nabla f(x) - \int_0^1 \nabla^2 f(x + t(\hat{x}-x))\, \nabla^2 f(x)^{-1} \nabla f(x)\, dt$$

$$= \left(\nabla^2 f(x) - \int_0^1 \nabla^2 f(x + t(\hat{x}-x))\, dt \right) \nabla^2 f(x)^{-1} \nabla f(x)$$

$$\nabla^2 f(x) \int_0^1 (1-tr)^2 dt \preceq \int_0^1 \nabla^2 f(x + t(\hat{x}-x))\, dt \preceq \nabla^2 f(x) \int_0^1 \frac{dt}{(1-tr)^2}$$

$$\nabla^2 f(x) \left(1 - r + \frac{r^2}{3}\right) \preceq \qquad\qquad \preceq \nabla^2 f(x) \cdot \frac{1}{1-r}$$

$$\therefore \left\| \nabla^2 f(x)^{-1/2} \left(\nabla^2 f(x) - \int_0^1 \nabla^2 f(x + t(\hat{x}-x))\, dt \right) \nabla^2 f(x)^{-1/2} \right\|_{op} \leq \max\left\{ \frac{r}{1-r}, r - \frac{r^2}{3} \right\}$$

$$= \frac{r}{1-r}.$$

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

$$= \left\| \nabla^2 f(x)^{-1/2} \left(\nabla^2 f(x) - \int_0^1 \nabla^2 f(x + t(\hat{x}-x))\, dt \right) \nabla^2 f(x)^{-1/2} \nabla^2 f(x)^{-1/2} \nabla f(x) \right\|_2$$

$$\leq \left\| \nabla^2 f(x)^{-1/2} \left(\qquad\qquad \right) \nabla^2 f(x)^{-1/2} \right\|_{op} \| \nabla^2 f(x)^{-1/2} \nabla f(x) \|_2$$

$$\leq \frac{r}{1-r} \cdot r = \frac{r^2}{1-r}. \qquad \blacksquare$$

Our goal is to reach $\nabla f(x) \to 0$.

We can measure progress in terms of $\| \nabla f(x) \|$.

We can measure progress in terms of ...

<u>Lemma 3.</u> Suppose $\|\nabla f(x)\|_{\nabla^2 f(x)^{-1}} \le \frac{1}{6}$. Then

$$\|x^* - x\|_x \le \frac{4}{3} \|\nabla f(x)\|_{\nabla^2 f(x)^{-1}}$$

$$\|x^* - x\|_{x^*} \le 2 \|\nabla f(x)\|_{\nabla^2 f(x)^{-1}}$$

$$f(x) - f(x^*) \le \|\nabla f(x)\|^2_{\nabla^2 f(x)^{-1}}$$

<u>Pf.</u> Let $r = \|x^* - x\|_x$.

$$\nabla f(x) = \underbrace{\nabla f(x^*)}_{(=0)} + \int_0^1 \nabla^2 f(x^* + t(x - x^*))(x - x^*)\,dt$$

$$\|\nabla f(x)\|_{\nabla^2 f(x)^{-1}} = \left\| \int_0^1 \nabla^2 f(x^* + t(x - x^*))(x - x^*)\,dt \right\|_{\nabla^2 f(x)^{-1}}$$

$$\ge \left\| \nabla^2 f(x)(x - x^*) \int_0^1 (1 - (1-t)r)^2\,dt \right\|_{\nabla^2 f(x)^{-1}}$$

$$= \int_0^1 (1 - (1-t)r)^2\,dt \; \|x - x^*\|_x = \left(1 - r + \frac{r^2}{3}\right) r \ge \frac{3}{4} r \quad \left(r \le \frac{1}{4}\right)$$

$$\Rightarrow r \le \frac{4}{3} \|\nabla f(x)\|_{\nabla^2 f(x)^{-1}} \le \frac{4}{3} \cdot \frac{1}{6} < \frac{1}{4}$$

$$\|x - x^*\|_{x^*} \le \frac{\|x - x^*\|_x}{1 - r} \le \frac{\frac{4}{3} \|\nabla f(x)\|_{\nabla^2 f(x)^{-1}}}{\left(1 - \frac{1}{4}\right)} < 2 \|\nabla f(x)\|_{\nabla^2 f(x)^{-1}}.$$

Finally, ... $\int_0^1$ ... $\nabla^T \nabla^2 f(x^* + t(x - x^*))(x - x^*)\,dt$

Finally

$$f(x) = f(x^*) + \langle \nabla f(x^*), x - x^* \rangle + \int_0^1 (1-t)(x-x^*)^T \nabla^2 f(x^* + t(x-x^*))(x-x^*)\, dt$$

$$f(x) - f(x^*) \le \|x - x^*\|_x^2 \int_0^1 \frac{1-t}{(1-(1-t)r)^2}\, dt$$

$$= \|x - x^*\|_x^2 \cdot \frac{1}{r^2}\left(\frac{r}{1-r} + \ln(1-r)\right)$$

$$\le \left(\frac{1}{2} + r\right) \|x - x^*\|_x^2 \le \|\nabla f(x)\|^2_{\nabla^2 f(x)^{-1}}.$$

IPM algo. sc barrier $\phi(x)$, ν

Start at $x_0 = \arg\min \phi(x)$.

$$\text{Min} \quad f_t(x) = t\, c^T x + \phi(x)$$

$$t_0 = \frac{1}{6} \|c\|^{-1}_{\nabla^2 \phi(x_0)^{-1}}$$

$$\text{While} \quad t \le \frac{\nu + \sqrt{\nu}}{\varepsilon}$$

$$\left[\begin{array}{l} x \leftarrow x - \nabla^2 f_t(x)^{-1} \nabla f_t(x) \\ t \leftarrow (1+h)\, t \end{array}\right. \qquad h = \frac{1}{9\sqrt{\nu}}.$$

Lemma $\phi : K \to \mathbb{R}$ ν-s.c.

$\forall x, y \in K$ $\langle \nabla\phi(x), y - x \rangle \le \nu$.

Pf Let $\alpha(t) = \langle \nabla\phi(x + t(y-x)), y - x \rangle$

$$\alpha'(t) = \|y-x\|^2_{x+t(y-x)}$$

$$|\alpha(t)| \le \|\nabla\phi(x+t(y-x))\|_{\nabla^2\phi(x+t(y-x))^{-1}} \|y-x\|_{x+t(y-x)}$$

$$\le \sqrt{\nu}\sqrt{\alpha'(t)}$$

$$\alpha'(t) \ge \frac{1}{\nu}\alpha(t)^2 \quad \text{i.e.} \quad \frac{d}{dt}\left(\frac{1}{\alpha(t)}\right) \le -\frac{1}{\nu}$$

$\alpha(0) \le 0$ ✓ else α is increasing $\Rightarrow \alpha(1) > 0$

$$0 \le \frac{1}{\alpha(1)} \le \frac{1}{\alpha(0)} - \frac{1}{\nu} \Rightarrow \alpha(0) \le \nu.$$

Lemma. (1) $c^T x_t \le c^T x^* + \frac{\nu}{t}$

(2) $\forall x$ with $\|tc + \nabla\phi(x)\|_{\nabla^2\phi(x)^{-1}} \le \frac{1}{6}$

$$c^T x \le c^T x^* + \frac{\nu + \sqrt{\nu}}{t}.$$

Pf. By optimality of x_t, $\nabla f_t(x_t) = tc + \nabla\phi(x_t) = 0$

Pf. By optimality of x_t, $\nabla f_t(x_t) = tc + \nabla\phi(x_t) = 0$

$$c^T(x_t - x^*) = -\frac{1}{t}\langle \nabla\phi(x_t), x_t - x^* \rangle \le \frac{\nu}{t}$$

For f_t, $\|x - x_t\|_x \le 2\|\nabla f_t(x)\|_{\nabla^2 f_t(x)^{-1}} \le \frac{1}{3}$.

$\nabla^2 f_t = \nabla^2\phi$

$$c^T(x - x_t) \le \|c\|_{\nabla^2\phi(x)^{-1}} \|x - x_t\|_x \le \frac{1}{3}\|x - x_t\|_x$$

using

$$c = \frac{\nabla f_t(x)}{t} - \frac{\nabla\phi(x)}{t}, \quad c^T(x - x_t) \le \frac{1}{t}\left(\|tc + \nabla\phi(x)\|_{\nabla^2\phi(x)^{-1}} + \|\nabla\phi\|_{\nabla^2\phi(x)^{-1}}\right) \cdot \|x - x_t\|_x$$

$$\le \frac{1}{3t}\left(\frac{1}{6} + \sqrt{\nu}\right) \le \frac{\sqrt{\nu}}{t}.$$

Thm. IPM finds $x \in K$ with $c^T x \le c^T x^* + \varepsilon$ in $O\left(\sqrt{\nu}\log\left(\frac{\nu}{\varepsilon}\|c\|_{\nabla^2\phi(x^0)^{-1}}\right)\right)$ iterations.

Pf. We maintain $\|\nabla f_t(x)\|_{\nabla^2 f_t(x)^{-1}} \le \frac{1}{6}$.

Then at $t = \frac{\nu + \sqrt{\nu}}{\varepsilon}$, we have the conclusion.

Given such an x, after one Newton step, we have

$$\|\nabla f_t(x)\| \le \left(\frac{1}{6}\right)^2 \le \frac{1}{\cdot}$$

given such that ... [illegible]

$$\|\nabla f_t(x)\|_{\nabla^2 f_t(x)^{-1}} \le \frac{\left(\frac{1}{6}\right)^2}{1-\frac{1}{6}} \le \frac{1}{25}.$$

Now $\hat{t} = (1+h)t$ $\quad f_{\hat{t}}(x) = (1+h)t\,C^T x + \phi(x)$

$$\nabla^2 f_{\hat{t}}(x) = \nabla^2 \phi(x)$$

$$\nabla f_{\hat{t}}(x) = (1+h)tC + \nabla\phi(x)$$

$$= (1+h)\nabla f_t(x) - h\nabla\phi(x)$$

So, $$\|\nabla f_{\hat{t}}(x)\|_{\nabla^2 f_{\hat{t}}(x)^{-1}} \le (1+h)\|\nabla f_t(x)\|_{\nabla^2 f_t(x)^{-1}} + h\|\nabla\phi(x)\|_{\nabla^2\phi(x)^{-1}}$$

$$\le \frac{1+h}{25} + h\sqrt{\nu} \le \frac{1+h}{25} + \frac{1}{9} \le \frac{1}{6}.$$