Lecture 12

Last time

- Complexity of Simplex
- Intro to interior point methods
- Remembering Newton.

Today

- Affine invariance
- A new guarantee for Newton's.

An affine invariant measure of progress

Last time, we recalled the Newton Method:

$$x_{k+1} \leftarrow x_k - [\nabla^2 f(x_k)]^{-1} \nabla f(x_k).$$

and presented

Theorem (*): Suppose that f is such that $\forall x$ near x^*

$$\alpha I \preceq \nabla f^2(x) \preceq \beta I$$

Then, for any point x_0 sufficiently close to $x^* = \operatorname{argmin} f$, we have

$$\|x_1 - x^*\|_2 \le \frac{\beta}{2\alpha} \|x_0 - x^*\|_2^2 .$$

$\dashv$

In particular we needed

$$\|x_0 - x^*\| < \frac{\alpha}{4\beta},$$

for quadratic convergence.

This property was terrible for our hopes of polynomial time guarantees for IPM because it depends on our representation A and how close we are to the boundary of the constraints.

Lemma: Consider $\varphi : \mathbb{R}^d \to \mathbb{R}^d$ given by

$Q \in \mathbb{R}^{d \times d}$ invertible

$$\varphi(x) = Qx + w$$

Let $\tilde{f} = f \circ \varphi$ and

$$x_1 = x_0 - [\nabla^2 f(x_0)] \nabla f(x_0),$$

$$y_1 = y_0 - [\nabla^2 \tilde{f}(y_0)]^{-1} \nabla \tilde{f}(y_0).$$

Then $y_1 = \varphi(x_1)$.

$\dashv$

Intuition

Thus, the iterates don't change under affine transformations and the convergence in "the right" metric should be independent of α and β.

A measure of progress should

1) Measure how far is $\nabla f(x_k)$ from being zero
2) Be affine invariant.

So, we might simply measure the norm of $\nabla f(x_k)$ in a different metric. Define

$$\langle u, v \rangle_x = u^T \nabla^2 f(x) v$$

$$\|u\|_x^2 = \langle u, u \rangle_x$$

Inner product if $\nabla^2 f(x) \succ 0$.

I claim that a good metric of progress is

$$\|n(x)\|_x = \|[\nabla^2 f(x)]^{-1} \nabla f(x)\|_x$$
$$= \left(\nabla f(x)^T [\nabla^2 f(x)]^{-1} \nabla f(x)\right)^{1/2}.$$

Exercise: check that $\|n(x)\|_x$ is zero iff $\nabla f(x) = 0$ and $\|n(x)\|_x$ is affine invariant.

An affine invariant guarantee

We now prove a guarantee in terms of $\|n(x)\|_x$. To this end we need a condition that ensures the continuity of $\|\cdot\|_x$.

Def: A C^2 function f is (strongly nondegenerate) self concordant (SC) if

$$(1 - 3\delta)\nabla^2 f(x) \preceq \nabla^2 f(y) \preceq (1+3\delta)\nabla^2 f(x)$$

for all $\|y - x\|_x = \delta < 1$. $\dashv$

Intuitively the Hessians change continuously.

Theorem: Suppose $f : \mathbb{R}^d \to \mathbb{R}$ is strictly convex and SC. Then, if $\|n(x_0)\|_{x_0} \leq \frac{1}{6}$ we have

$$\|n(x_1)\|_{x_1} \leq 3\,\|n(x_0)\|_{x_0}^2 .$$

independent of conditioning

Before we prove this result let's introduce the shorthand $H(x) = \nabla^2 f(x)$. Further, for a given $Q \succ 0$ we use

$$\|u\|_Q^2 = u^T Q u .$$

Proof: Note that $\|n(x)\|_x = \|\nabla f(x)\|_{H(x)^{-1}}$. Since

$$\|x_1 - x_0\|_{x_0} = \|n(x_0)\|_{x_0} \leq 1/6,$$

and f is SC, we have

$$\frac{1}{2} H(x_0) \preceq H(x_1) \preceq 2 H(x_0).$$

In turn, this implies (Why?)

$$\frac{1}{2} H(x_0)^{-1} \preceq H(x_1)^{-1} \preceq 2 H(x_0)^{-1}!$$

Therefore

$$\begin{aligned}\|n(x_1)\|_{x_1} &= \nabla f(x_1)^T H(x_1)^{-1} \nabla f(x_1) \\ &\le 2 \; \nabla f(x_1) H(x_0)^{-1} \nabla f(x_1) \\ &= 2 \; \|\nabla f(x_1)\|_{H(x_0)^{-1}}\end{aligned}$$

Thus, it suffices to show

$$\|\nabla f(x_1)\|_{H(x_0)^{-1}} \le \frac{3}{2} \|\nabla f(x_0)\|^2_{H(x_0)^{-1}}.$$

By the Fundamental Theorem of calculus

$$\begin{aligned}&\nabla f(x_1) \\ &= \nabla f(x_0) + \int_0^1 H(x_0 + t(x_1 - x_0))(x_1 - x_0)\, dt \\ &= \nabla f(x_0) - \int_0^1 H(x_0 + t(x_1 - x_0)) H(x_0)^{-1} \nabla f(x_0)\, dt \\ &= \nabla f(x_0) - \left[\int_0^1 H(x_0 + t(x_1 - x_0))\, dt\right] H(x_0)^{-1} \nabla f(x_0) \\ &= \underbrace{\left[H(x_0) - \int_0^1 H(x_0 + t(x_1 - x_0))\, dt\right]}_{M(x_0)} H(x_0)^{-1} \nabla f(x_0)\end{aligned}$$

$$= M(x_0) H(x_0)^{-1} \nabla f(x_0).$$

Thus,

$$\|\nabla f(x_1)\|_{H(x_0)^{-1}}$$

$$= \| M(x_0) H(x_0)^{-1} \nabla f(x_0) \|_{H(x_0)^{-1}}$$

(why?)

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

$$\le \| H(x_0)^{-1/2} M(x_0) H(x_0)^{-1/2} \|_{op} \| H(x_0)^{-1/2} \nabla f(x_0) \|_2$$

$$= \| H(x_0)^{-1/2} M(x_0) H(x_0)^{-1/2} \|_{op} \| \nabla f(x_0) \|_{H(x_0)^{-1}}.$$

Cauchy-Schwarz

The result would follow if we show

$$\| H(x_0)^{-1/2} M(x_0) H(x_0)^{-1/2} \|_{op} \le \frac{3}{2} \| \nabla f(x_0) \|_{H(x_0)^{-1}}$$

For this we use

Lemma: Suppose $A \in S^n_+$ and $B \in S^n$ such that for some $\alpha > 0$

$$-\alpha A \preceq B \preceq \alpha A.$$

Then

$$\|A^{-1/2} B A^{-1/2}\|_{op} \le \alpha.$$

Proof of the Lemma: By definition

$$\|A^{-1/2} B A^{-1/2}\|_{op} = \sup_{u \neq 0} \frac{|u^T A^{-1/2} B A^{-1/2} u|}{u^T u}$$

Change of variables $v \leftarrow A^{-1/2} u$ $= \sup_{v \neq 0} \frac{|v^T B v|}{v^T A v}$

By assumption $\le \alpha.$

$\square$

Thus, we need to show

(**) $$-\frac{3}{2} \delta H(x_0) \preceq M(x_0) \preceq \frac{3}{2} \delta H(x_0).$$

with $\delta = \|\nabla f(x_0)\|_{H(x_0)^{-1}}$. Since f is SC we that for $\delta = \|x_1 - x_0\|_{x_0}$ and for all $t \in [0,1]$

$$-3t\delta H(x_0) \preceq H(x_0) - H(x_0 + t(x_1 - x_0))$$
$$\preceq 3t\delta H(x_0).$$

Integrating from $t=0$ to $t=1$ yields (**), which proves the result. $\square$

Q: But is this applicable in our setting?

Lemma: The log barrier function

$$B(x) = -\sum_{i=1}^{m} \log(\underbrace{b_i - a_i^T x}_{r_i(x)})$$

is self concordant.

Proof: Recall that

$$H(x) = \sum_{i=1}^{m} \frac{a_i a_i^T}{s_i(x)^2}$$

Let $\delta = \|y - x\|_x < 1$, then

$$\delta^2 = (y-x)^T H(x)(y-x) = \sum_{i=1}^{m} \left(\frac{a_i^T(y-x)}{s_i(x)}\right)^2.$$

Hence, each individual term

$$\left(\frac{s_i(y) - s_i(x)}{s_i(x)}\right)^2 = \left(\frac{a_i^T(y-x)}{s_i(x)}\right)^2 \leq \delta^2.$$

Therefore,

$$(1-\delta)|s_i(x)| \leq |s_i(y)| \leq (1+\delta)|s_i(x)|$$

which implies

$$\frac{(1+\delta)^{-2}}{s_i(x)^2} \le \frac{1}{s_i(y)^2} \le \frac{(1-\delta)^{-2}}{s_i(x)^2}.$$

Thus

$$\frac{(1+\delta)^{-2}}{s_i(x)^2} a_i a_i^T \preceq \frac{a_i a_i^T}{s_i(y)^2} \preceq \frac{(1-\delta)^{-2}}{s_i(x)^2} a_i a_i^T.$$

Summing over all i and using the fact that

$$(1-3\delta) \le (1+\delta)^{-2} \le (1-\delta)^{-2} \le 1+3\delta$$

for all $\delta \in [0,1]$, yields the result. $\square$