Lecture #16

Last Time: Gradient Descent

(what types of fctns? smooth, strongly convex)

Today: Interior Point Methods

A new approach for convex optimization; following Karmarkar, Nesterov, Nemirovskii

$$\min c^T x \quad \text{s.t. } x \in \underset{\text{convex}}{K}$$

is to work with

$$(*) \quad \min \underbrace{t}_{\text{scalar}} c^T x + \underbrace{F(x)}_{\text{barrier}}$$

where $F(x) \xrightarrow[x \to \partial K]{} \infty$, instead.

Example: Let $K = \{x \mid Ax \leq b\}$, then we can choose $F(x) = -\sum_{i=1}^{m} \log(b_i - (Ax)_i)$

Now let $x^*(t)$ be the argmin to $(*)$. The set:

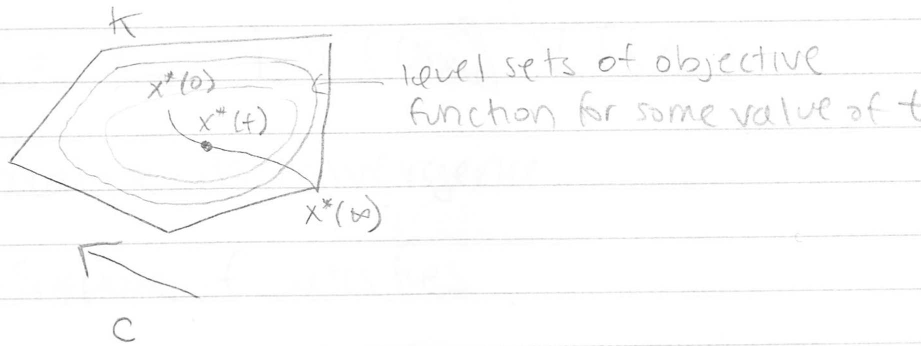
$$\{x^*(t) \mid 0 \leq t \leq \infty\}$$

is called the central path and smoothly interpolates between

$$x^*(0) = \text{"analytic center"}$$

and optimal soln to original problem

Let's draw a figure:



Main Ingredients:

(1) Newton's Method: If we are close to $x^*(t)$, converge quadratically to it

(2) From $x^*(t)$, can consider $t' > t$ and converge quickly to $x^*(t')$

(3) Fast methods to find $x^*(0)$ (won't cover)

Let's start with the traditional analysis of Newton's Method

Setup: Let $f: \mathbb{R}^n \to \mathbb{R}$ be C^2. Then

$$f(x+h) = f(x) + h^T \nabla f(x) + \frac{1}{2} h^T \nabla^2 f(x) h + o(\|h\|^2)$$

Then $\underset{h}{\operatorname{argmin}}\ h^T \nabla f(x) + \frac{1}{2} h^T \nabla^2 f(x) h =$

$$-[\nabla^2 f(x)]^{-1} \nabla f(x)$$

assume nonnegative eigenvalues

Now we can define Newton's method (in high-dimensions)

Newton's Method:

$$X_{k+1} = X_k - [\nabla^2 f(X_k)]^{-1} \nabla f(X_k)$$

Next we analyze its convergence:

Theorem 1: Suppose f satisfies

(1) $\|\nabla^2 f(x) - \nabla^2 f(y)\|_{op} \leq R \|x-y\|$

(2) $\nabla^2 f(x^*) \succeq \mu I$ (i.e. all eigenvalues of $\nabla^2 f(x^*)$ are at least μ)

x^* a local minimum of f

(3) $\|x_0 - x^*\| \leq \frac{\mu}{2R}$

Then $\|x_{k+1} - x^*\| \leq \frac{R}{\mu} \|x_k - x^*\|^2$ (quadratic convergence)

Proof: We use the following elemetary formula:

$$\nabla f(x+h) - \nabla f(x) = \int_0^1 \nabla^2 f(x+sh)\, h\, ds$$

$$\Rightarrow \nabla f(x_k) - \overset{0}{\cancel{\nabla f(x^*)}} = \underbrace{\int_0^1 \nabla^2 f(x^* + s(x_k - x^*))(x_k - x^*)\, ds}$$

Now we can write

$$x_{k+1} - x^* = \underbrace{x_k - x^*}_{\nabla^2 f(x_k)^{-1}(\nabla^2 f(x_k)(x_k - x^*))} - (\nabla^2 f(x_k))^{-1} \nabla f(x_k)$$

Thus we have

$$x_{k+1} - x^* = \nabla^2 f(x_k)^{-1} \underbrace{\int_0^1 \left(\nabla^2 f(x_k) - \nabla^2 f(x^* + s(x_k - x^*))\right)(x_k - x^*)\,ds}_{\leq sR\|x_k - x^*\|^2}$$ (i.e. bdd in norm)

$$\Rightarrow \|x_{k+1} - x^*\| \leq \left(\underbrace{\|(\nabla^2 f(x_k))^{-1}\|_{op}}_{\overset{\text{claim}}{\leq} \frac{2}{\mu}}\right)\left(\frac{R}{2}\|x_k - x^*\|^2\right)$$

Now $\lambda_{min}(\nabla^2 f(x_k)) \geq \underbrace{\lambda_{min}(\nabla^2 f(x^*))}_{\geq \mu} - \lambda_{max}\underbrace{(\nabla^2 f(x^*) - \nabla^2 f(x_k))}_{\leq R\|x_k - x^*\| \leq \frac{\mu}{2}}$

Thus $\|(\nabla^2 f(x))^{-1}\|_{op} \leq \frac{2}{\mu}$ $\boxtimes$

---

what if we apply the transformation

$y = Ax$? (here A is invertible)

Let $\phi(y) = f(A^{-1}y) = f(x)$. what does Newton's method do?

$$\nabla \phi(y) = (A^{-1})^T \nabla f(A^{-1}y)$$

$$\nabla^2 \phi(y) = (A^{-1})^T \nabla^2 f(A^{-1}y) A^{-1}$$

Thus $\left[\nabla^2 \phi(y)\right]^{-1} \nabla \phi(y) =$

$$A\left(\nabla^2 f(A^{-1}y)\right)^{-1} \nabla f(A^{-1}y)$$

which is the same as computing steps in x-space and transforming back to y.

"Newton's method is affine invariant"

although guarantees of Theorem 1 are not.

definition: [Newton Decrement]

$$\lambda_f(x) := \sqrt{\nabla f(x)^T \left(\nabla^2 f(x)\right)^{-1} \nabla f(x)}$$ affine invariant

Then one can show:

Theorem 2: Let f satisfy

just need it on $x \in K$ $\rightarrow$ $\nabla^3 f(x)[h,h,h] \leq 2\left(h^T \nabla^2 f(x) h\right)^{3/2}$ self concordance

Then $$\lambda_f\Big(\underbrace{x - \left(\nabla^2 f(x)\right)^{-1}\nabla f(x)}_{\text{newton step}}\Big) \leq \left(\frac{\lambda_f(x)}{1-\lambda_f(x)}\right)^2$$

This gives quadratic converge in an affine invariant way, for $\{x \mid \lambda_f(x) < 1\} \equiv$ "Newton decrement ball"

Now we are ready to proceed with Step (2).

We want to understand how large we can set t' s.t.

$$\lambda_{F_{t'}}(x^*(t)) < \frac{1}{4}$$

where $F_{t'}(x) = t' c^T x + F(x)$. Now:

$$\nabla F_{t'}(x^*(t)) = \overset{0}{\cancel{\nabla F_t(x^*(t))}} + (t'-t)c$$

Thus $\lambda_{F_{t'}}(x^*(t)) = (t'-t)\sqrt{c^T(\underbrace{\nabla^2 F_{t'}(x^*(t))}_{\nabla^2 F(x^*(t))})^{-1}c}$

For a natural class of functions

$$\nabla^2 F(x) \succeq \frac{1}{\nu} \nabla F(x) \nabla F(x)^T$$

Then we have $\nabla F_t(x^*(t)) = 0 = tc + \nabla F(x^*(t))$ so $c = \frac{-\nabla F(x^*(t))}{t}$

$$\lambda_{F_{t'}}(x^*(t)) \le \frac{(t'-t)\sqrt{\nu}}{t}$$

And now suppose we want:

$$\frac{(t'-t)\sqrt{\nu}}{t} < \frac{1}{4} \iff t' \le t\left(1 + \frac{1}{4\sqrt{\nu}}\right)$$

It is now not hard to show that to get an ε-approximate solution we need

i.e. boost from t_0 to ν/ε

$$k = O\left(\sqrt{\nu} \log \frac{\nu}{t_0 \varepsilon}\right) \text{ iterations}$$

Collecting the conditions we've used:

(a) $\nabla^3 F(x)[h,h,h] \le 2\left(h^T \nabla^2 F(x) h\right)^{3/2}$

(b) $\nabla^2 F(x) \succeq \frac{1}{\nu} \nabla F(x) \nabla F(x)^T$

we call F ν-self concordant $\Rightarrow$ interior point methods

Moreover one can show that the log-barrier satisfies these conditions for

$$\nu = m$$

where A is $m \times n$.