yilmaz

The Interior Point Method.

Can we do this for general convex optimization?

$$\min_{x \in K} f(x) \qquad f, K \text{ convex}.$$

Reduce to

$$\boxed{\begin{array}{l} \min t \\ x \in K \\ f(x) \le t \end{array}} \qquad \begin{array}{l} \min c^T x \\ x \in K. \end{array}$$

Previously K: $x_i \geq 0 \quad Ax = b$

or $\quad Ax \le b$.

For $\min c^T x$

$Ax \le b$

we could have used $\quad \min \; c^T x - t \sum_i \ln(b_i - A_i x)$

and proceeded as before.

alternatively $\quad \min \; t c^T x \underbrace{- \sum \ln(b_i - A_i x)}_{\phi(x).}$

t goes from $0 \to \infty$.

$\phi(x)$ is convex.

blows up at the boundary.

So keeps x in the feasible region.

Can we do this in general? How?

$$\min_{x \in K} c^T x \longrightarrow \min \phi_t(x) = t c^T x + \phi(x)$$

$\phi : K \to \mathbb{R}$ convex

$x \to \partial K \quad \phi(x) \to \infty$.

$$x_t = \arg\min \phi_t(x).$$

IPM

- start with x close to CP for some $t_0 > 0$.
- while t not large enough
  - (step) move to x closer to x_t
  - $t \leftarrow (1+h)t$.

How to initialize? How to step? How to set h?

We will choose ϕ to make these easy and efficient.

Easy: GD. best case convergence is "linear" $\sim \log \frac{1}{\varepsilon}$.
not affine invariant.

E.g. $f(x,y) = 100x^2 + y^2$

GD: $(x,y) \leftarrow (x,y) - h\nabla f(x,y) = (1-200h)x,\ (1-2h)y$.

We need $h \leq \frac{1}{L} = \frac{1}{200}$ making it slow.

Newton method: At x go to

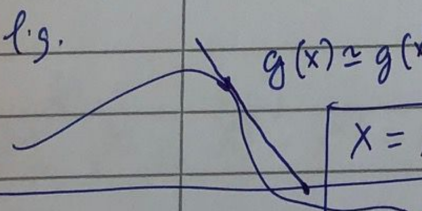
$$\operatorname*{argmin}_y \ f(x) + (y-x)^T \nabla f(x) + \frac{1}{2}(y-x)^T \nabla^2 f(x)(y-x)$$

$$\nabla f(x) + \nabla^2 f(x)(y-x) = 0$$

$$\boxed{y = x - (\nabla^2 f(x))^{-1} \nabla f(x)}$$

Can be used to find roots

e.g.



$$g(x) \simeq g(x^{(k)}) + g'(x^{(k)})(x - x^{(k)}) = 0$$

$$\boxed{x = x^{(k)} - \frac{g(x^{(k)})}{g'(x^{(k)})}}$$

$$g = \nabla f(x) \rightarrow \quad \nabla f(x) \approx \nabla f(x^{(k)}) + \nabla^2 f(x^{(k)})(x - x^{(k)})$$

$$x^{(k+1)} = x^{(k)} - \nabla^2 f(x^{(k)})^{-1} \nabla f(x^{(k)}).$$

On the example $\nabla f(x,y) = \begin{pmatrix} 200x \\ 2y \end{pmatrix}$ $\quad \nabla^2 f(x,y) = \begin{pmatrix} 200 & 0 \\ 0 & 2 \end{pmatrix}$ $\quad \nabla^2 f(x)^{-1} \nabla f(x) = \begin{pmatrix} \frac{1}{200} & 0 \\ 0 & \frac{1}{2} \end{pmatrix} \begin{pmatrix} 200x \\ 2y \end{pmatrix} = \begin{pmatrix} x \\ y \end{pmatrix}$

✓

Lema. Newton iteration is affine-invariant.

Pf.

$$g(y) = f(Ay)$$

$$\nabla g(y) = A^T f(Ay)$$

$$\nabla^2 g(y) = A^T \nabla^2 f(Ay) A$$

$$\nabla^2 g(y)^{-1} \nabla g(y) = (A^T \nabla^2 f(Ay) A)^{-1} A^T \nabla f(Ay) = A^{-1} \nabla^2 f(x)^{-1} \nabla f(x).$$

$$\underline{y - \nabla^2 g(y)^{-1} \nabla g(y) = A^{-1}(x - \nabla^2 f(x)^{-1} \nabla f(x))}$$

Newton iteration gives "quadratic" convergence when close enough to fixed point.

Thm. Suppose $|f'(x^*)| \geq \alpha$ at $f(x^*) = 0$, f' is L-Lipschitz. $f: \mathbb{R} \to \mathbb{R}$. Then for $|x^{(0)} - x^*| \leq \frac{\alpha}{2L}$, $|x^{(k)} - x^*| \leq \frac{\alpha}{L}\left(\frac{L}{\alpha}|x - x^*|\right)^2$

error is squaring! $\sim \log\log\frac{1}{\varepsilon}$ iterations.

Pf.

$$x^{(k+1)} - x^* = x^{(k)} - x^* - \frac{f(x^{(k)})}{f'(x^{(k)})}$$

$$0 = f(x^*) = f(x^{(k)}) + \int_{x^{(k)}}^{x^*} f'(z)\,dz = f(x^{(k)}) + f'(x^{(k)})(x^* - x^{(k)}) + \int_{x^{(k)}}^{x^*} (f'(z) - f'(x^{(k)}))\,dz$$

So

$$x^{(k+1)} - x^{(*)} = x^{(k)} - x^{(*)} - \frac{f(x^{(k)})}{f'(x^{(k)})} = \frac{1}{f'(x^{(k)})} \int_{x^{(k)}}^{x^*} (f'(z) - f'(x^{(k)}))\,dz \leq \frac{L}{f'(x^{(k)})} \int_{x^{(k)}}^{x^*} |z - x^{(k)}|\,dz$$

$$= \frac{L}{2f'(x^{(k)})} |x^* - x^{(k)}|^2.$$

Since $f'(x^{(k)}) \geq f'(x^*) - L|x^{(k)} - x^*|$

$$|x^{(k+1)} - x^*| \leq \frac{L}{2\left(\alpha - L \cdot \frac{\alpha}{2L}\right)} |x^{(k)} - x^*|^2 = \frac{L}{\alpha}|x^{(k)} - x^*|^2 < \frac{1}{2}|x^{(k)} - x^*|$$

So inductively $|x^{(k)} - x^*| < \frac{\alpha}{2L}$. $\frac{L}{\alpha}|x^{(k+1)} - x^*| \leq \left(\frac{L}{\alpha}|x^{(k)} - x^*|\right)^2$.

A quick application.

Thm. Let $g(x)$ be a real rooted polynomial $g(x) = a\prod_{i=1}^{n}(x-\lambda_i)$

Given $\lambda > \lambda_1$ $\qquad \lambda_n \le \lambda_{n-1} \le \dots \le \lambda_1$

Newton iteration finds x s.t. $\lambda_1 \le x \le \lambda_1 + \varepsilon(\lambda - \lambda_1)$

in $O(n \log \frac{1}{\varepsilon})$ iterations.

Pf. $g'(x) = a_n \sum_i \prod_{j \ne i} (x-\lambda_j)$ $\qquad \frac{g(x)}{g'(x)} = \frac{1}{\sum_i \frac{1}{x-\lambda_i}}$

So $\frac{g(x)}{g'(x)} \le \frac{1}{\frac{1}{x-\lambda_1}} = x - \lambda_1$

$\Rightarrow 0 \le x^{(k+1)} - \lambda_1 \le (1-\frac{1}{n})(x^{(k)} - \lambda_1)$.

and $\ge \frac{1}{n \cdot \frac{1}{x-\lambda_1}} = \frac{x-\lambda_1}{n}$

$\Rightarrow O(n \log \frac{1}{\varepsilon})$ iterations.

We will choose $\phi(x)$ in IPM so that Newton iteration works well.

$x \in \mathbb{R}^n$

$f: \mathbb{R}^n \to \mathbb{R}$ convex. $\qquad \|v\|_x^2 = v^T \nabla^2 f(x) v \qquad \|v\|_{\nabla^2 f(x)} \qquad Df = \nabla f$

Def 1. f is self-concordant: $\forall h \in \mathbb{R}^n$ $\left|D^3 f(x)[h,h,h]\right| \le 2\|h\|_x^3$

$h_1, h_2, h_3 \in \mathbb{R}^n$ $\left|D^3 f(x)[h_1,h_2,h_3]\right| \le 2\|h_1\|_x \|h_2\|_x \|h_3\|_x$

Eg. $f(x) = -\ln x$

$Df(x) = f'(x) = -\frac{1}{x} \qquad D^2 f(x) = \frac{1}{x^2} \qquad D^3 f(x) = -\frac{2}{x^3}$.

Equivalently, on any line. $g(t) = f(x+th)$ satisfies

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

Def: $\phi : K \to \mathbb{R}$ is a ν-self-concordant barrier if

ϕ is convex

$\phi(x) \to \infty$ as $x \to \partial K$.

$\|\nabla \phi(x)\|^2_{\nabla^2 \phi(x)^{-1}} \le \nu$.

Example. $\left| \frac{f'(x)^2}{f''(x)} \right| = \frac{\frac{1}{x^2}}{\frac{1}{x^2}} = 1$. $-\ln(x)$ is 1-self-concordant.

$\phi(x) = -\sum_{i=1}^{n} \ln x_i \qquad \nabla \phi(x) = \begin{pmatrix} -\frac{1}{x_i} \\ \vdots \end{pmatrix} \qquad \nabla^2 \phi(x) = \begin{pmatrix} \frac{1}{x_i^2} & & 0 \\ & \ddots & \\ 0 & & \end{pmatrix}$

$\nabla \phi(x)^T \nabla^2 \phi(x)^{-1} \nabla \phi(x) = n$. Alternatives $|D\phi(x)[h]| \le \sqrt{\nu}\left(D^2\phi(x)[h,h]\right)^{\frac{1}{2}}$

Lemma. For self-concordant f, with $\hat{x} = x - \nabla^2 f(x)^{-1} \nabla f(x)$

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

(b) else $\hat{x} \leftarrow x - \frac{1}{(1 + \|\nabla f(x)\|_{\nabla^2 f(x)^{-1}})} \nabla^2 f(x)^{-1} \nabla f(x)$

then $f(\hat{x}) \le f(x) - \|\nabla f(\hat{x})\|_{\nabla^2 f(\hat{x})^{-1}} + \ln(1 + \|\nabla f(\hat{x})\|_{\nabla^2 f(\hat{x})^{-1}})$.

iterations is $O\left(f(x^{(0)}) - \min f + \ln\ln \frac{1}{\varepsilon}\right)$ to reach $\|\nabla f(x)\|_{\nabla^2 f(x)^{-1}} \le \varepsilon$.

If $\underbrace{\|\nabla \phi(x)\|_{\nabla^2 \phi(x)^{-1}}}_{=\lambda} \le 1$, then $f(x) - \min f(x) \le -\lambda - \ln(1-\lambda)$

$\phi(x)$ has barrier parameter ν

$\frac{1}{2}\phi(x)$ ———————— $\frac{\nu}{2}$.

Thm (barrier calculus).

$g = f(Ay+b)$ preserves self-concordance.

$\phi_1(x) + \phi_2(x)$ is $\nu_1 + \nu_2$ s.c. on $K_1 \cap K_2$.

$\phi_1(x) + \phi_2(x)$ is $\nu_1 + \nu_2$ on $K_1 \times K_2$.