APPM 5630 Optim., spring '21 — Stephen Becker

Running out of time, condense Newton + IPM to 1 lecture

... previously, I motivated why we need a new analysis for Newton.

recall: **Def** $f:\mathbb{R}\to\mathbb{R}$ is self-concordant if convex

and $|f'''(x)| \le 2\cdot(f''(x))^{3/2}$

Ex: $-\log(x)$, all quadratics, all linear functions

Def $f:\mathbb{R}^n\to\mathbb{R}$ is s.c. if $\forall x,v\in\mathbb{R}^n$, $t\mapsto f(x+tv)$ is s.c.

Fact s.c. is affine invariant (because of 3/2 exponent)

Ex: $-\log\det(X)$

Fact: $\begin{pmatrix}\nabla^2 f(x)\succ 0\\ f \text{ s.c.}\end{pmatrix} \iff \begin{pmatrix} f \text{ strictly cvx}\\ f\text{ s.c.}\end{pmatrix}$ ... so we'll assume this...

Damped / Guarded Newton [suitable if f strictly cvx]

Repeat:

$\Delta x_{nt} = -(\nabla^2 f_k)^{-1}\cdot\nabla f_k$, $\quad \nabla f_k := \nabla f(x_k)$ etc.

$\|x\|_H := \sqrt{\langle x | H | x\rangle}$ $\quad \lambda_k^2 = \lambda(x_k) = \underset{\text{"Newton decrement"}}{\|\Delta x_{nt}\|_{\nabla^2 f}} = \sqrt{\langle\nabla f_k | \nabla^2 f_k^{-1} | \nabla f_k\rangle}$

terminate if $\lambda_k^2/2 \le \varepsilon$

backtrack linesearch, starting at $t=1$,

(ensure: $f(x_k + t\cdot\Delta x_{nt}) < f(x_k) + \alpha\cdot t\cdot\underbrace{\langle\nabla f_k, \Delta x_{nt}\rangle}_{=\lambda_k^2}$) for some $0<\alpha<.5$

$x_{k+1} = x_k + t\cdot\Delta x_{nt}$

let $p^* = \min f(x)$

Fact (BV04, §9.6.3)

If $\lambda(x) < 0.68$, then $f(x) - p^* \le \lambda(x)^2$ so valid stopping crit.

Thm (under assumptions, eg strict convexity) — bad if far... too good if close...

$\exists\, 0<\eta<1/4$, $\gamma>0$, s.t.

$k=1,\dots,k_0$ (I) Damped Newton phase: $\lambda(x_k)>\eta$, $f(x_{k+1}) - f(x_k) < -\gamma$

$k>k_0$ (II) Quad. Conv. phase: $\lambda(x_k)\le\eta$, $t=1$, $2\lambda_{k+1} \le (2\cdot\lambda_k)^2$

i.e. $f(x_{k_0+l}) - p^* \le 1/4\cdot\left(\tfrac{1}{2}\right)^{2^{l-k_0+1}}$, $\quad 10^{-10}$ $l=5$ iter, 10^{-20} $l=6$ iter, 10^{-40} $l=7$ iter

... So, # iter to reach ε-sol'n is

$$\underbrace{\frac{f(x_0)-p^*}{\gamma}}_{\text{Phase 1}} + \underbrace{\log(\log(1/\varepsilon))}_{\text{Phase 2}} \qquad \text{or w/ practical \#'s, } \gamma \geq 1/375$$

$$\approx 375\cdot(f(x_0)-p^*) + 6$$ since $\log(\log(1/\varepsilon)) < 6$ for all conceivable $\varepsilon > 0$.

- EQUALITY CONSTRAINTS $\quad \min f(x)$ s.t. $Ax=b$

KKT: $\nabla f(x) + A^T\nu = 0 \longrightarrow$ linearize: $\quad \Delta x = x - x_k$

$\underbrace{Ax = b}$ $\qquad \nabla f(x) \approx \nabla f_k + \nabla^2 f_k \cdot \Delta x$

$A(x_k + \Delta x) = b$

So... solve $\begin{bmatrix} \nabla^2 f_k & A^T \\ A & 0 \end{bmatrix} \cdot \begin{bmatrix} \Delta x \\ \nu \end{bmatrix} = \begin{bmatrix} -\nabla f_k \\ b - A\cdot x_k \end{bmatrix}$ — $= 0$ if x_k already feasible (if not, adjust linesearch)

"KKT system"

Equiv., change variables: $Ax = b \iff x = Fz + x_p$ $\quad (Ax_p = b)$, $AF = 0$, $R(F) = N(A)$

Solve $\min_z f(Fz + x_p)$.

Either way, it's still basically Newton: solve lin. eq. each step

See book for details, issues

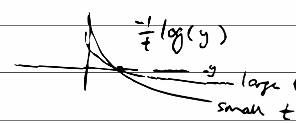
- INEQUALITY AND EQUALITY CONSTRAINTS:

Interior-Point Methods (IPM)

(P) $\quad \min f_0(x)$
$\quad$ s.t. $f_i(x) \leq 0$, $i=1,\dots,m$ $\}$ convex
$\quad Ax = b$ $\leftarrow$ Easy to deal with
$\quad$ ($p \times n$)

assume: f_i convex, $C^2(\mathbb{R}^n)$,
$\exists$ optimal sol'n, + strict feasibility.

Barrier method: $\quad -\frac{1}{t}\cdot\log(y)$



solve

$$\min_{Ax=b} f_0(x) + \underbrace{\sum_{i=1}^m \frac{-1}{t}\log(-f_i(x))}_{\varphi_t(x)}$$, convex and often s.c. (eg, f_i = affine)

$t \to \infty$, but solve sequence of $t_k \nearrow \infty$ for faster convergence ("central path", $\{x^*(t)\}_{t>0}$

$$x^* = \operatorname*{argmin}_{Ax=b} f_0 + \varphi_t$$

Look at KKT optimality conditions for a point $x^*(t)$ on the central path:

$$(KKT)_t \quad \cdot\ 0 = \overbrace{\nabla f(x) + \sum_{i=1}^{m} \frac{-1}{t \cdot f_i(x)} \cdot \nabla f_i(x)}^{\nabla \varphi_t(x)} + A^T \nu \quad \Bigg] \text{ at } x = x^*(t)$$

$\cdot\ Ax = b$, and note $f_i(x) < 0$ naturally via log

Compare to original problem's KKT

- $0 = \nabla f_0(x) + \sum \lambda_i \nabla f_i(x) + A^T \nu$ since $\mathcal{L}(x,\lambda,\nu) = f_0(x) + \sum \lambda_i f_i(x) + \nu^T(Ax-b)$
- $Ax = b$, $f_i(x) \leq 0$
- $\lambda \geq 0$
- $-\lambda_i f_i(x) = 0$, $i = 1, \dots, m$

From $(KKT)_t$, define $\lambda_i = \dfrac{-1}{t \cdot f_i(x)}$, same ν ← at $x = x^*(t)$

plug into dual fcn for original problem

$$g(\lambda,\nu) = \min_x \mathcal{L}(x,\lambda,\nu)$$

$$= \mathcal{L}(x^*(t), \lambda, \nu) \text{ via stationarity condition of } (KKT)_t$$

$$= f_0(x^*(t)) + \sum_{i=1}^{m} \underbrace{\frac{-1}{t \cdot \cancel{f_i(x^*(t))}}}_{\lambda_i} \cdot \cancel{f_i(x^*(t))} + \nu^T \underbrace{(Ax^*(t) - b)}_{=0}$$

$$= f_0(x^*(t)) - \frac{m}{t}$$

So via weak duality, $g(\lambda,\nu) \leq p^*$

i.e., $f_0(x^*(t)) - m/t \leq p^*$, so $\boxed{f(x^*(t)) - p^* \leq m/t}$

So... t controls accuracy: Choose t large

(but large t requires many Newton steps, so solve for a sequence $t_k \nearrow \infty$, e.g. $t_k = 5 \cdot t_{k-1}$, $\underbrace{\text{warm-starting}}_{\text{"path-following"}}$)

Goes back to Fiacco & McCormick '60s

much analysis, improvements in 80's, 90's another derivation: take complementary slackness condition from KKT eq'n for original problem:

$-\lambda_i \cdot f_i(x) = 0$. Relax to $-\lambda_i \cdot f_i(x) = \frac{1}{t}$.

Central path

Details / Issues:

- we don't find $x^*(t)$ exactly, only up to error in Newton's method...

  Sol'n #1: no big deal, Newton's method is super accurate

  sol'n #2: just take 1 step of Newton, since no need to "over-optimize" when t is small.

  Very effective: "primal-dual" IPM. See §11.7

- How to find strictly feasible starting pt.?

  Solve pre-processing or "Phase I" problem:

$$p^* = \min_{\substack{x \in \mathbb{R}^n \\ s \in \mathbb{R}}} \; s \quad f_i(x) \le s, \; Ax = b$$

This itself has an easy strictly feasible pt: any x_0, and $s_0 > \max_i f_i(x_0)$

If $p^* < 0$, great, we have strictly feas. pt.

If $p^* > 0$, (P) is infeasible

If $p^* = 0$, it's infeasible (if min not obtained) or not strictly feasible so IPM won't work

Analysis: see §11.5. Via self-concordancy, you can analyze total complexity! ($\underbrace{\text{Inner}}_{\text{Newton}}$ + $\underbrace{\text{outr}}_{t_k}$ steps)

Take-aways

- works well for generalized inequalities,

  i.e., $f_i(x) \le 0$ is $f_i(x) \succeq_K 0$ for some nice cone K

  and $f_i(x) = c_i^T x - d_i$

... as long as the cone has a s.c. barrier

Ex: $K = \mathbb{R}^n_+$, $-\log(x)$ is a s.c. barrier

Ex: $K = S^n_+$ (PSD matrices), $-\log\det(X)$ is one

... So, due to high-accuracy, IPM common choice for conic problems (LP, convex QP, SOCP, SDP) in $\underbrace{\text{moderate dimensions}}_{\text{don't scale well due to cost of Newton step (solving lin. eq.)}}$.

- Many variants, eg., Mehotra predictor-corrector method initialization tricks, special considerations for SDPs

  see Steve Wright's '97 "IPM for LP" book