Lecture 13
Convergence analysis of the barrier method

- complexity analysis of the barrier method
  - convergence analysis of Newton's method
  - choice of update parameter μ
  - bound on the total number of Newton iterations
- initialization

Complexity analysis

we'll analyze the method of page 12–21 with

- update $t^+ = \mu t$
- starting point $x^*(t^{(0)})$ on the central path

main result: #Newton iters is bounded by

$$O(\sqrt{m}\log(\epsilon^{(0)}/\epsilon)) \qquad \text{(where } \epsilon^{(0)} = m/t^{(0)}\text{)}$$

caveats:

- methods with good worst-case complexity don't necessarily work better in practice
- we're not interested in the numerical values for the bound—only in the exponent of m and n
- doesn't include initialization
- insights obtained from analysis are more valuable than the bound itself

Outline

1. convergence analysis of Newton's method for

$$\varphi(x) = tc^T x - \sum_{i=1}^{m} \log(b_i - a_i^T x)$$

   (will give us a bound on the number of Newton steps per outer iteration)

2. effect of μ on total number of Newton iterations to compute $x^*(\mu t)$ from $x^*(t)$

3. combine 1 and 2 to obtain the total number of Newton steps, starting at $x^*(t^{(0)})$

The Newton decrement

Newton step at x:

$$\begin{aligned} v &= -\nabla^2 \varphi(x)^{-1} \nabla \varphi(x) \\ &= -(A^T \mathbf{diag}(d)^2 A)^{-1} (tc + A^T d) \end{aligned}$$

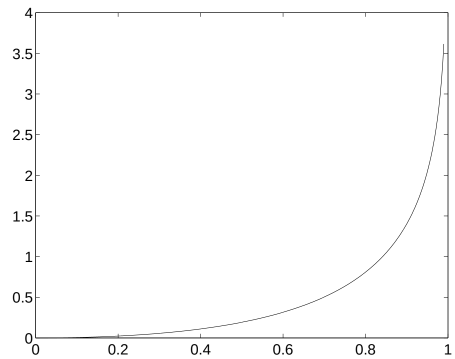
where $d = (1/(b_1 - a_1^T x), \ldots, 1/(b_m - a_m^T x))$

Newton decrement at x:

$$\begin{aligned} \lambda(x) &= \sqrt{\nabla \varphi(x)^T \nabla^2 \varphi(x)^{-1} \nabla \varphi(x)} \\ &= \sqrt{v^T \nabla^2 \varphi(x) v} \\ &= \left(\sum_{i=1}^{m} \left(\frac{a_i^T v}{b_i - a_i^T x} \right)^2 \right)^{1/2} \\ &= \|\mathbf{diag}(d) A v\| \end{aligned}$$

theorem. if $\lambda = \lambda(x) < 1$, then φ is bounded below and

$$\varphi(x) \leq \varphi(x^*(t)) - \lambda - \log(1-\lambda)$$



- if $\lambda \leq 0.81$, then $\varphi(x) \leq \varphi(x^*(t)) + \lambda$
- useful as stopping criterion for Newton's method

proof: w.l.o.g. assume $b - Ax = \mathbf{1}$; let $x^* = x^*(t)$, $z = \mathbf{1} + Av$

$$\lambda = \|Av\| < 1 \quad \Longrightarrow \quad z = \mathbf{1} + Av \geq 0$$

$$\nabla^2\varphi(x)v = A^TAv = -\nabla\varphi(x) = -tc - A^T\mathbf{1} \quad \Longrightarrow \quad A^Tz = -tc$$

$$\begin{aligned}
tc^Tx^* - \sum_{i=1}^m \log(b_i - a_i^Tx^*) &= -z^TAx^* - \sum_{i=1}^m \log(b_i - a_i^Tx^*) \\
&\geq -z^TAx^* + \sum_{i=1}^m \log z_i - z^T(b - Ax^*) + m \\
&= -(\mathbf{1} + Ax)^Tz + \sum_{i=1}^m \log z_i + m \\
&= tc^Tx + \sum_i (-a_i^Tv + \log(1 + a_i^Tv)) \\
&\geq tc^Tx + \lambda + \log(1-\lambda)
\end{aligned}$$

inequalities follow from:

1. $\log y \leq -\log z + zy - 1$ for $y, z > 0$

2. $\sum_{i=1}^{m}(y_i - \log(1 + y_i)) \leq -\|y\| - \log(1 - \|y\|)$ if $\|y\| < 1$

Local convergence analysis

$$x^+ = x - \nabla^2\varphi(x)^{-1}\nabla\varphi(x)$$

theorem: if $\lambda < 1$, then $Ax^+ < b$ and $\lambda^+ \leq \lambda^2$

(λ is Newton decrement at x; λ^+ is Newton decrement at x^+)

- gives bound on number of iterations: suppose we start at $x^{(0)}$ with $\lambda^{(0)} \leq 0.5$, then $\varphi(x) - \varphi(x^*(t)) < \delta$ after fewer than

$$\log_2 \log_2(1/\delta) \text{ iterations}$$

- called region of **quadratic convergence**
- practical rule of thumb: 5–6 iterations

proof.

1. $\lambda^2 = \sum_{i=1}^m (a_i^T v)^2/(b_i - a_i^T x)^2 < 1$ implies $a_i^T(x+v) < b_i$

2. assume $b - Ax^+ = \mathbf{1}$; let $w = \mathbf{1} - d - \mathbf{diag}(d)^2 Av$

$$\begin{aligned}
(\lambda^+)^2 = \|Av^+\|^2 &= \|Av^+\|^2 - 2(Av^+)^T(w + Av^+) & (1)\\
&\leq \|w + Av^+ - Av^+\|^2 \\
&= \sum_{i=1}^m (1-d_i)^4 & (2)\\
&= \sum_{i=1}^m (d_i a_i^T v)^4 \\
&\leq \|\mathbf{diag}(d)Av\|^4 = \lambda^4
\end{aligned}$$

(1) uses $A^T w = tc + A^T\mathbf{1}$, $A^T A v^+ = -tc - A^T\mathbf{1}$
(2) uses $Av = Ax^+ - b - Ax + b = -\mathbf{1} + \mathbf{diag}(d)^{-1}\mathbf{1}$,
therefore $d_i a_i^T v = 1 - d_i$ and $w_i = (1-d_i)^2$

Global analysis of Newton's method

damped Newton algorithm: $x^+ = x + sv$, $v = -\nabla^2\varphi(x)^{-1}\nabla\varphi(x)$

step size to the boundary: $s = \alpha^{-1}$ where

$$\alpha = \max\left\{ \frac{a_i^T v}{b_i - a_i^T x} \;\middle|\; a_i^T v > 0 \right\} \qquad (\alpha = 0 \text{ if } Av \leq 0)$$

theorem. for $s = 1/(1+\alpha)$,

$$\varphi(x+sv) \leq \varphi(x) - (\lambda - \log(1+\lambda))$$

- very simple expression for step size
- same bound if s is determined by an exact line search

if $\lambda \geq 0.5$,

$$\varphi(x + (1+\alpha)^{-1}v) \leq \varphi(x) - 0.09$$

(hence, convergence)

proof. define $f(s) = \varphi(x+sv)$ for $0 \le s < 1/\alpha$

$$f'(s) = v^T \nabla \varphi(x+sv), \qquad f''(s) = v^T \nabla^2 \varphi(x+sv)^T v$$

for Newton direction v: $f'(0) = -f''(0) = -\lambda^2$

by integrating the upper bound

$$f''(s) = \sum_{i=1}^m \left(\frac{a_i^T v}{b_i - a_i^T x - s a_i^T v} \right)^2 \le \frac{f''(0)}{(1-s\alpha)^2}$$

twice, we obtain

$$f(s) \le f(0) + s f'(0) - \frac{f''(0)}{\alpha^2} (s\alpha + \log(1-s\alpha))$$

upper bound is minimized by $s = -f'(0)/(f''(0) - \alpha f'(0)) = 1/(1+\alpha)$

$$\begin{aligned} f(s) &\le f(0) - \frac{f''(0)}{\alpha^2}(\alpha - \log(1+\alpha)) \\ &\le f(0) - (\lambda - \log(1+\lambda)) \qquad \text{(since } \alpha \le \lambda\text{)} \end{aligned}$$

Summary

given x with $Ax < b$, tolerance $\delta \in (0, 0.5)$
repeat

1. *Compute Newton step at x:* $v = -\nabla^2 \varphi(x)^{-1} \nabla \varphi(x)$
2. *Compute Newton decrement:* $\lambda = (v^T \nabla^2 \varphi(x) v)^{1/2}$
3. If $\lambda \le \delta$, return(x)
4. *Update x:* If $\lambda \ge 0.5$,
   $x := x + (1+\alpha)^{-1} v$ where $\alpha = \max\{0, \max_i a_i^T v/(b_i - a_i^T x)\}$
   else, $x := x + v$

upper bound on #iterations, starting at x:

$$\log_2 \log_2 (1/\delta) + 11\,(\varphi(x) - \varphi(x^*(t)))$$

usually very pessimistic; good measure in practice:

$$\beta_0 + \beta_1\,(\varphi(x) - \varphi(x^*(t)))$$

with empirically determined β_i ($\beta_0 \le 5$, $\beta_1 \ll 11$)

#Newton steps per outer iteration

#Newton steps to minimize $\varphi(x) = t^+c^Tx - \sum_{i=1}^m \log(b_i - a_i^Tx)$

theorem. if $z > 0$, $A^Tz + c = 0$, then

$$\varphi(x^*(t^+)) \geq -t^+b^Tz + \sum_{i=1}^m \log z_i + m(1 + \log t^+)$$

in particular, for $t^+ = \mu t$, $z_i = 1/t(b_i - a_i^Tx^*(t))$:

$$\varphi(x^*(t^+)) \geq \varphi(x^*(t)) - m(\mu - 1 - \log \mu)$$

yields estimates for #Newton steps to minimize φ starting at $x^*(t)$:

$$\beta_0 + \beta_1 m(\mu - 1 - \log \mu)$$

- is an upper bound for $\beta_0 = \log_2 \log_2(1/\delta)$, $\beta_1 = 11$
- is a good measure in practice for empirically determined β_0, β_1

proof. if $z > 0$, $A^Tz + c = 0$, then

$$\begin{aligned}
\varphi(x) &= t^+c^Tx - \sum_{i=1}^m \log(b_i - a_i^Tx) \\
&\geq t^+c^Tx + \sum_{i=1}^m \log z_i - t^+z^T(b - Ax) + m(1 + \log t^+) \\
&= -t^+b^Tz + \sum_{i=1}^m \log z_i + m(1 + \log t^+)
\end{aligned}$$

for $z_i = 1/(t(b_i - a_i^Tx^*(t))$, $t^+ = \mu t$, this yields

$$\varphi(x^\star(t^+)) \geq \varphi(x^\star(t)) - m(\mu - 1 - \log \mu)$$

Bound on total #Newton iters

suppose we start on central path with $t = t^{(0)}$

number of outer iterations:

$$\#\text{outer iters} = \left\lceil \frac{\log(\epsilon^{(0)}/\epsilon)}{\log \mu} \right\rceil$$

- $\epsilon^{(0)} = m/t^{(0)}$: initial duality gap
- $\epsilon^{(0)}/\epsilon$: reduction in duality gap

upper bound on total #Newton steps:

$$\left\lceil \frac{\log(\epsilon^{(0)}/\epsilon)}{\log \mu} \right\rceil (\beta_0 + \beta_1 m(\mu - 1 - \log \mu))$$

- $\beta_0 = \log_2 \log_2(1/\delta)$, $\beta_1 = 11$
- can use empirical values for β_i to estimate average-case behavior

Strategies for choosing μ

- μ independent of m:

$$\#\text{Newton steps per outer iter} \le O(m)$$

$$\text{total }\#\text{Newton steps} \le O(m \log(\epsilon^{(0)}/\epsilon))$$

- $\mu = 1 + \gamma/\sqrt{m}$ with γ independent of m

$$\#\text{Newton steps per outer iter} \le O(1)$$

$$\text{total }\#\text{Newton steps} \le O(\sqrt{m} \log(\epsilon^{(0)}/\epsilon))$$

  follows from:
  - $m(\mu - 1 - \log \mu) \le \gamma^2/2$, because $x - x^2/2 \le \log(1+x)$ for $x > 0$
  - $\log(1 + \gamma/\sqrt{m}) \ge \log(1+\gamma)/\sqrt{m}$ for $m \ge 1$

Choice of initial t

rule of thumb: given estimate $\widehat{p}$ of $p^\star$, choose

$$m/t \approx c^T x - \widehat{p}$$

(since m/t is duality gap)

via complexity theory (c.f. page 13–12) given dual feasible z, #Newton steps in first iteration is bounded by an affine function of

$$\begin{aligned} & t(c^T x + b^T z) + \phi(x) - \sum_{i=1}^m \log z_i - m(1 + logt) \\ = \quad & t(c^T x + b^T z) - m \log t + \text{const.} \end{aligned}$$

choose t to minimize bound; yields $m/t = c^T x + b^T z$

there are many other ways to choose t