Lecture 8:

Introduction to Sequential Unconstrained Minimization for Constrained Optimization

- brief history of SUMT & IP methods
- logarithmic barrier function
- central path
- basic SUMT

History of SUMT & IP methods

Interior point methods (very roughly)

- smooth 'barrier' function replaces constraints
- solve sequence of smooth unconstrained problems

Early methods (1950s–1960s)

- Frisch, SUMT (Fiacco & McCormick), Dikin, method of centers (Huard & Lieu)
- convergence theory, but no worst-case complexity theory
- (often) worked well in practice
- fell out of favor in 1970s

New methods (1984–)

- initiated by Karmarkar (for LP)
- polynomial worst-case complexity
- work well in practice
- extended to general case by Nesterov & Nemirovsky 1988

Logarithmic barrier function

consider smooth, convex problem
(for now, without equality constraints):

$$\begin{array}{ll} \text{minimize} & f_0(x) \\ \text{subject to} & f_i(x) \le 0, \quad i = 1, \ldots, m \end{array}$$

assume strict feasibility:

$$C = \{x \mid f_i(x) < 0, \ i = 1, \ldots, m\} \neq \emptyset$$

we define *logarithmic barrier* ϕ as

$$\phi(x) = \begin{cases} -\sum_{i=1}^{m} \log(-f_i(x)) & x \in C \\ +\infty & \text{otherwise} \end{cases}$$

- ϕ is convex, smooth on C
- $\phi \to \infty$ as x approaches boundary of C

$\operatorname{argmin} \phi$ (if it exists) is called *analytic center* of inequalities $f_1(x) < 0, \ldots, f_m(x) < 0$

Central path

for $t \geq 0$ define

$$x^*(t) = \operatorname{argmin}(tf_0(x) + \phi(x))$$

(we assume minimizer exists and is unique)

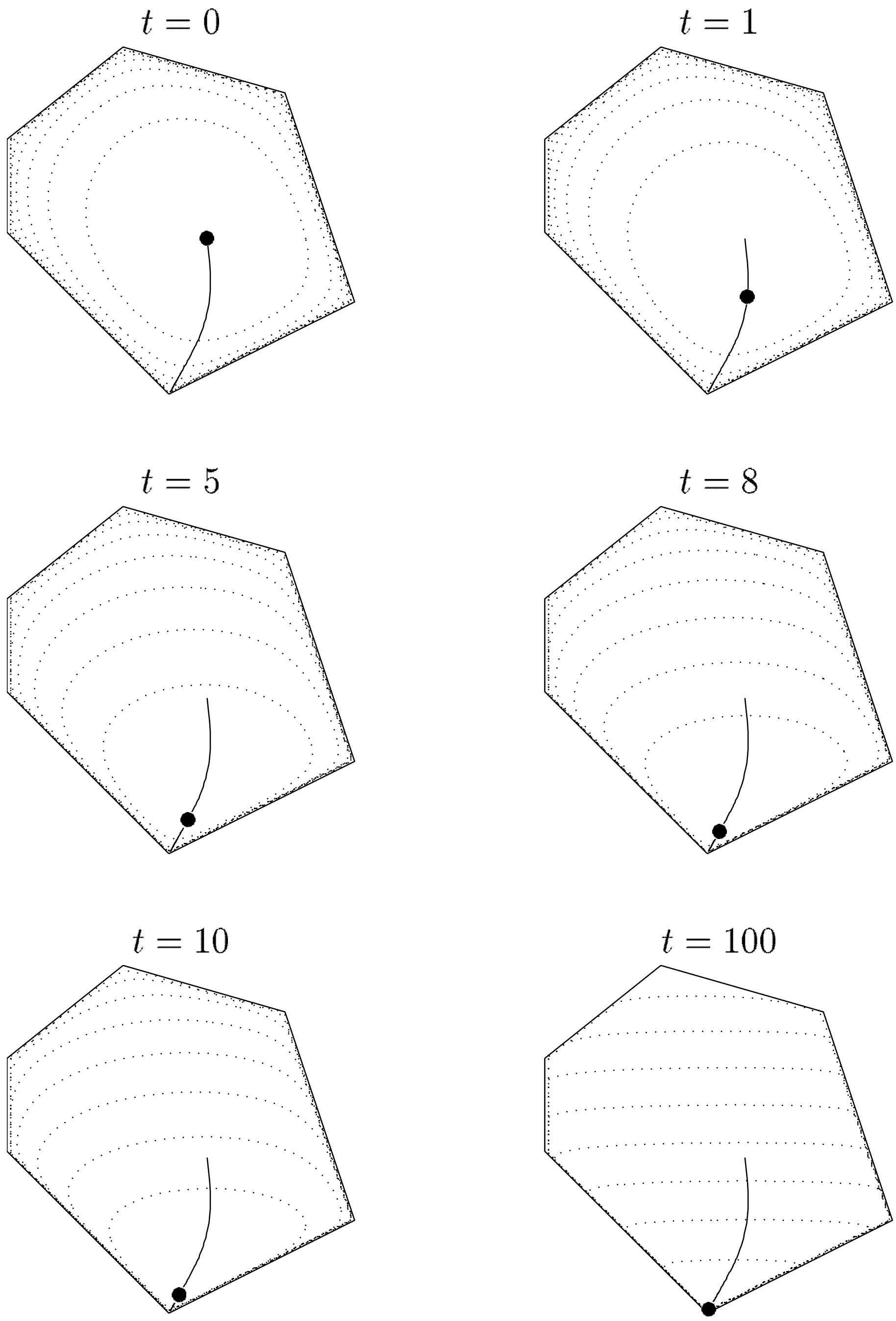
- curve $x^*(t)$ for $t \geq 0$ called *central path*
- can compute $x^*(t)$ by solving smooth effectively unconstrained minimization problem
  (given a strictly feasible starting point)
- t gives relative weight of objective and barrier
- barrier 'traps' $x^*(t)$ in strictly feasible set
- intuition suggests $x^*(t)$ converges to optimal as $t \to \infty$

$x^*(t)$ characterized by

$$t\nabla f_0(x^*(t)) + \sum_{i=1}^{m} \frac{1}{-f_i(x^*(t))} \nabla f_i(x^*(t)) = 0$$

Example: central path for LP

$x \in \mathbf{R}^2$, $A \in \mathbf{R}^{6\times 2}$, c points up



Force field interpretation

imagine a particle in C, subject to forces

ith constraint generates *force field*

$$F_i(x) = \nabla\left(-\log(-f_i(x))\right) = \frac{1}{-f_i(x)}\nabla f_i(x)$$

- ϕ is *potential* associated with constraint forces
- constraint forces push particle away from boundary of feasible set
- constraint forces trap particle in C

superimpose *objective force field*

$$F_0(x) = -t\nabla f_0(x)$$

- pulls particle toward small f_0
- t scales objective force

at $x^*(t)$, constraint forces exactly balance objective force

as t increases, particle is pulled towards optimal point, trapped in C by barrier potential

Central points and duality

recall $x^* = x^*(t)$ satisfies

$$t\nabla f_0(x^*) + \sum_{i=1}^m \frac{1}{-f_i(x^*)}\nabla f_i(x^*) = 0$$

rewrite as:

$$\nabla f_0(x^*) + \sum_{i=1}^m \lambda_i \nabla f_i(x^*) = 0, \qquad \lambda_i = \frac{1}{-f_i(x^*)t} > 0$$

so x^* also minimizes $L(x,\lambda) = f_0(x) + \sum_{i=1}^m \lambda_i f_i(x)$

i.e., λ is dual feasible and

$$\begin{aligned} f^\star &\geq g(\lambda) = \inf_x \left(f_0(x) + \sum_i \lambda_i f_i(x) \right) \\ &= f_0(x^*) + \sum_i \lambda_i f_i(x^*) \\ &= f_0(x^*) - m/t \end{aligned}$$

summary: a point on central path yields dual feasible point and lower bound:

$$f_0(x^*(t)) \geq f^\star \geq f_0(x^*(t)) - m/t$$

(which proves $x^*(t)$ becomes optimal as $t \to \infty$)

Central path and KKT conditions

KKT optimality conditions: x optimal $\Longleftrightarrow \exists\lambda$ s.t.

$$\begin{aligned} f_i(x) &\leq 0 \\ \lambda_i &\geq 0 \\ \nabla f_0(x) + \sum_i \lambda_i \nabla f_i(x) &= 0 \\ \lambda_i f_i(x) &= 0 \end{aligned}$$

centrality conditions: x central $\Longleftrightarrow \exists\lambda$, $t \geq 0$ s.t.

$$\begin{aligned} f_i(x) &\leq 0 \\ \lambda_i &\geq 0 \\ \nabla f_0(x) + \sum_i \lambda_i \nabla f_i(x) &= 0 \\ \lambda_i f_i(x) &= -1/t \end{aligned}$$

- for t large, $x^*(t)$ ‘almost’ satisfies KKT
- central path is continuous deformation of KKT condition

Unconstrained minimization method

given strictly feasible x, desired accuracy $\epsilon > 0$

1. $t := m/\epsilon$
2. compute $x^*(t)$ starting from x
3. $x := x^*(t)$

- computes ϵ-suboptimal point on central path (and certificate λ)
- solves constrained problem by solving one (effectively) unconstrained minimization (via Newton, BFGS, ...)
- works, but can be slow

SUMT

(Sequential Unconstrained Minimization Technique)

given strictly feasible x, $t > 0$, tolerance $\epsilon > 0$
repeat
1. compute $x^*(t)$ starting from x
2. $x := x^*(t)$
3. if $m/t \leq \epsilon$, return(x)
4. increase t

- generates sequence of points on central path
- solves constrained problem via sequence of unconstrained minimizations (often, Newton)
- simple updating rule for t: $t^+ = \mu t$ (typical values $\mu \approx 10 \sim 100$)

steps 1–4 above called **outer iteration**

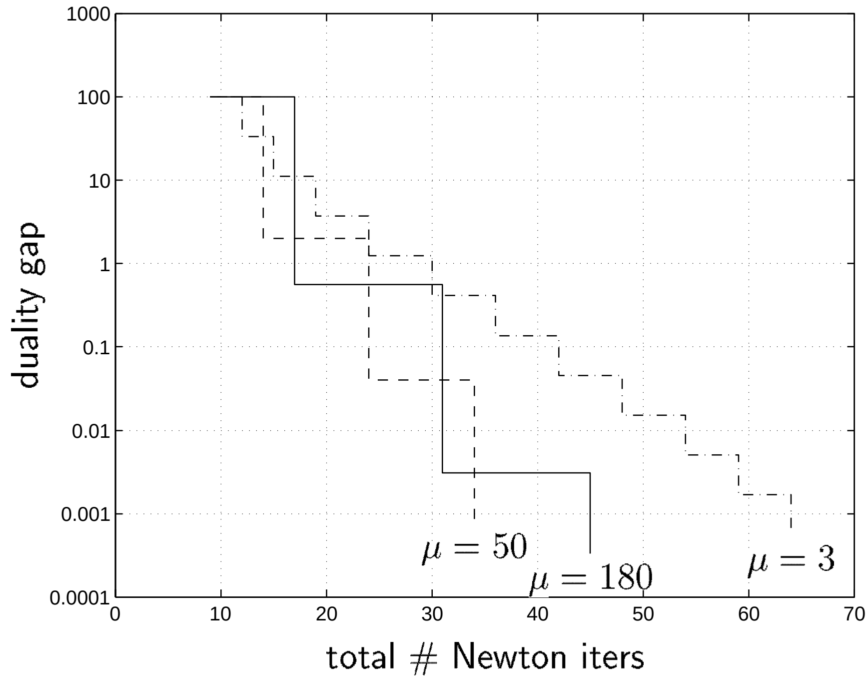
step 1 involves **inner iterations** (*e.g.*, Newton steps)

tradeoff: small $\mu \Longrightarrow$ few inner iters to compute $x^{(k+1)}$ from $x^{(k)}$, but more outer iters

Example: LP

$$\begin{array}{ll} \text{minimize} & c^T x \\ \text{subject to} & Ax \preceq b \end{array}$$

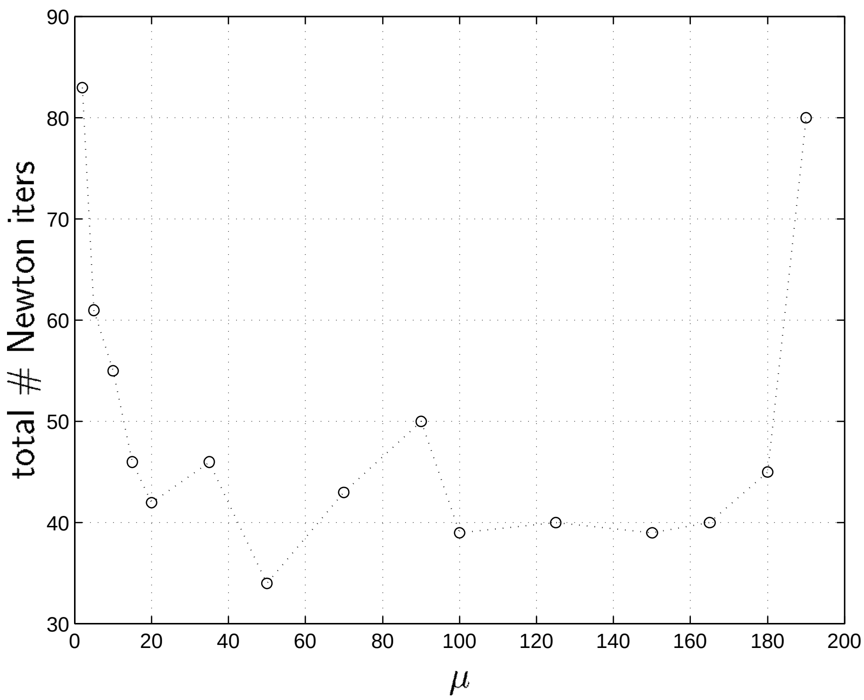
$A \in \mathbf{R}^{100 \times 50}$, Newton with exact line search



- width of 'steps' shows # Newton steps per outer iteration
- height of 'steps' shows reduction in duality gap $(1/\mu)$
- problem solved ($i.e.$, gap reduced by 10^5) in few tens of Newton iters
- gap decreases geometrically
- can see trade-off in choice of μ

LP example continued . . .

trade-off in choice of μ: $\#$ Newton iters required to reduce duality gap by 10^6



- SUMT works very well for wide range of μ