# Lecture 17
# Sequential unconstrained minimization

- brief history of SUMT & IP methods
- logarithmic barrier function
- central path
- UMT & SUMT
- complexity analysis
- feasibility phase
- generalized inequalities



## 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

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

$f_i$ convex, differentiable
(no equality constraints for simplicity)

assume strict feasibility:

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

define **logarithmic barrier** $\phi$ as

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

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

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



# Central path

$$x^*(t) = \textbf{argmin}(tf_0(x) + \phi(x)) \text{ for } t > 0$$

(we assume minimizer exists and is unique)

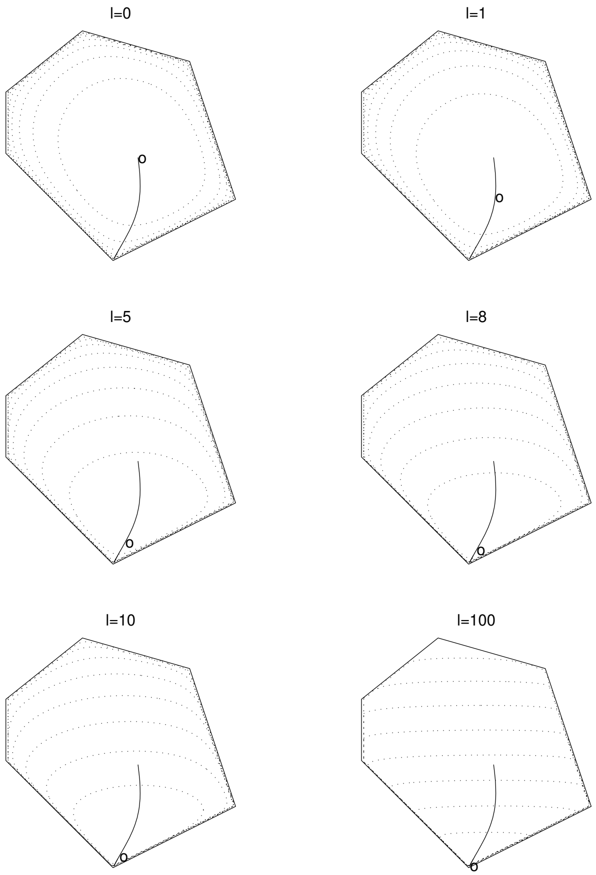
- curve $x^*(t)$ for $t \geq 0$ called **central path**
- can compute $x^*(t)$ by solving smooth unconstrained minimization problem
  (given a strictly feasible starting point)
- $t$ gives relative weight of objective and barrier
- barrier 'traps' $x^*(t)$ in strictly feasible set
- intuition suggests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

$i$th constraint generates **constraint force field**

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

- $\phi$ 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 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

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

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

*i.e.*, $\lambda$ is dual feasible and

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

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

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

(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.

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

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

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

– 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 $\epsilon$-suboptimal point on central path (and certificate $\lambda$)
- solves constrained problem by solving one smooth 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**

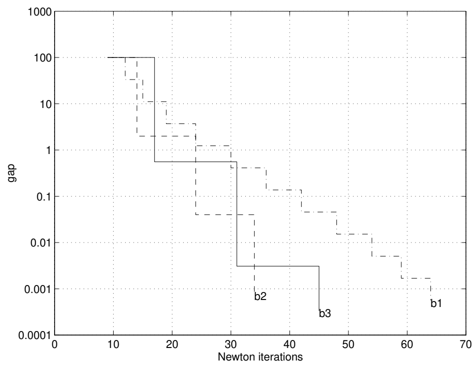
step 1 involves **inner iterations** (*e.g.*, Newton steps)

**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

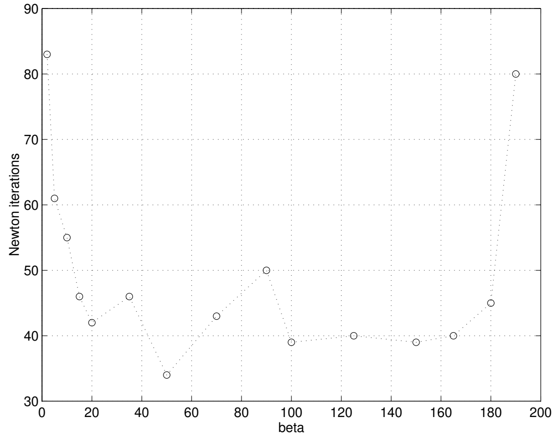



– width of 'steps' shows #Nt. iters per outer iter.
– height of 'steps' shows reduction in dual. gap ($1/\mu$)
– gap reduced by $10^5$ in few tens of Newton iters
– gap decreases geometrically
– can see trade-off in choice of $\mu$



LP example continued ...

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




SUMT works very well for wide range of $\mu$

## Complexity analysis

analyze tradeoff for $\mu$ using convergence theory of Newton's method

**number of outer iterations** (starting at $x^\star(t^{(0)})$):

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

**number of Newton steps per outer iteration:**

– starting at $x^\star(t)$, compute $x^\star(\mu t)$
– self-concordance assumption:

$$tf_0(x) + \phi(x)$$

is self-concordant for all $t \geq t^{(0)}$
– bound on #Newton steps:

$$c + \frac{\mu t f_0(x^\star(t)) + \phi(x^\star(t)) - \textbf{inf}_z(\mu t f_0(z) + \phi(z))}{\eta_2}$$

($c$, $\eta_2$ don't depend on problem parameters)



**examples** where $tf_0 + \phi$ is self-concordant

– LP, QCQP

$$\begin{array}{ll} \text{minimize} & x^T A_0 x + 2b_0^T x + c_0 \\ \text{subject to} & x^T A_i x + 2b_i^T x + c_i \leq 0, \quad i = 1, \ldots, m \end{array}$$

– entropy maximization

$$\begin{array}{ll} \text{minimize} & \sum_{i=1}^n x_i \log x_i \\ \text{subject to} & a_i^T x \leq b_i, \quad i = 1, \ldots, m \\ & x \succeq 0 \end{array}$$

($D = \{ x \mid x \succ 0 \}$)

– SOCP

$$\begin{array}{ll} \text{minimize} & c^T x \\ \text{subject to} & \dfrac{\|A_i x + b_i\|^2}{c_i^T x + d_i} \leq c_i^T x + d_i, \quad i = 1, \ldots, m \\ & c_i^T x + d_i \geq 0, \quad i = 1, \ldots, m \end{array}$$

($D = \{ x \mid c_i^T x + d_i > 0, \quad i = 1, \ldots, m \}$)

# Explicit bound on # Newton steps

**result from duality:** if $\lambda \succ 0$, then

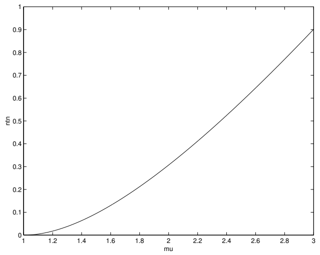
$$\inf_z \left(tf_0(z) + \phi(z)\right) \geq tg(\lambda) + \sum_{i=1}^m \log \lambda_i + m(1 + \log t)$$

in particular, $\lambda = \lambda^\star(t)$ yields:

$$\mu t f_0(x^\star(t)) + \phi(x^\star(t)) - \inf_z \left(\mu t f_0(z) + \phi(z)\right)$$
$$\leq m(\mu - 1 - \log\mu)$$

**upper bound on #Newton steps/outer iteration**:

$$\#\text{Newton steps} \leq c + \frac{m(\mu - 1 - \log\mu)}{\eta_2}$$

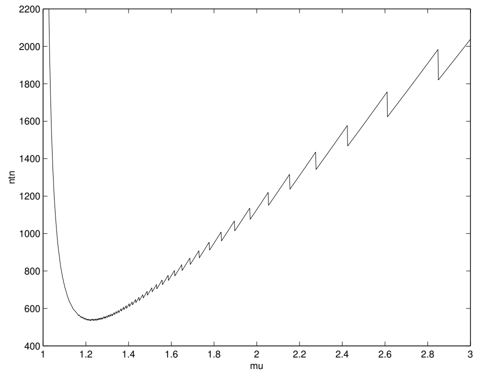




# Bound on total # Newton iters

**upper bound on total #Newton steps**:

$$\left\lceil \frac{\log(m/t^{(0)}\epsilon)}{\log\mu} \right\rceil \left(c + \frac{m(\mu - 1 - \log\mu)}{\eta_2}\right)$$




($c = 5$, $\eta_2 = 1/20$, $m = 10$, $m/t^{(0)}\epsilon = 10^5$)

- confirms trade-off in choice of $\mu$
- optimal $\mu$ depends on $m$, $\eta_2$, $c$, $t^{(0)}$, $\epsilon$

could use empirical values for $\eta_2$, $c$ to optimize average-case behavior (instead of worst-case)

## Strategies for choosing $\mu$ (vs. $m$)

- $\mu$ independent of $m$:

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

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

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

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

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



## Choice of initial $t$

**rule of thumb:** given estimate $\hat{f}$ of $f^\star$, choose

$$m/t \approx f_0(x) - \hat{f}$$

(since $m/t$ is duality gap)

**via complexity theory** ($tf_0 + \phi$ S.C.): given dual feasible $\lambda$, #Newton steps in first iteration is bounded by (affine function of)

$$\begin{aligned} & tf_0(x) + \phi(x) - \min_z(tf_0(z) + \phi(z)) \\ \leq\ & tf_0(x) + \phi(x) - t\tilde{g}(\lambda) - \sum_i \log \lambda_i - m(1 + \log t) \\ =\ & t(f_0(x) - g(\lambda)) - m \log t + \text{const.} \end{aligned}$$

choose $t$ to minimize bound; yields

$$m/t = f_0(x) - g(\lambda)$$

(same as rule of thumb)

there are many other ways to choose $t$

# Phase I

to compute strictly feasible point (or determine none exists) set up auxiliary problem:

$$\begin{array}{ll} \text{minimize} & w \\ \text{subject to} & f_i(x) \leq w, \ i = 1, \ldots, m \end{array}$$

- easy to find strictly feasible initial point (hence SUMT can be used)
- can use stopping criterion with target value $0$

if we include constraint on $f_0(x)$,

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

central path yields dual feasible point of original problem (hence, initial $t$)



# Generalized inequalities

standard problem with generalized inequalities:

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

- $f_0 : \mathbf{R}^n \to \mathbf{R}$ is convex, differentiable
- $f_i : \mathbf{R}^n \to \mathbf{R}^{m_i}$ are $K_i$-convex, differentiable

$\psi$ is a **log barrier** for cone $K \subseteq \mathbf{R}^m$ if

- $\mathbf{dom}\,\psi = \mathbf{Int}K$
- $\psi$ is convex and $K$-increasing
- there is a $\theta$ s.t. for all $a > 0$, $z \succeq_K 0$,

$$\psi(az) = \psi(z) - \theta \log a$$

generalizes logarithm from $\mathbf{R}_+$ to cone $K$

**example.** $\psi(Z) = \log\det Z^{-1}$ is a log barrier for PSD cone $K \subseteq \mathbf{R}^{n\times n}$, with $\theta = n$

# Central path with generalized ineq.

**logarithmic barrier** for constraints:

$$\phi(x) = \begin{cases} \sum_{i=1}^{L} \psi_i(-f_i(x)) & \text{if } f_i(x) \preceq_{K_i} 0,\ i = 1, \ldots, L \\ \infty & \text{otherwise} \end{cases}$$

**central path:** $x^*(t) = \mathbf{argmin}_x (t f_0(x) + \phi(x))$

characterized by

$$t\nabla f_0(x) + \nabla \phi(x) = 0$$

$$t\nabla f_0(x) - \sum_{i=1}^{L} D f_i(x)^T \nabla \psi_i(-f_i(x)) = 0$$



# Central path and duality

$x^*(t)$ yields dual feasible point

$$\lambda_i^*(t) = -\frac{1}{t} \nabla \psi_i(-f_i(x^*(t)))$$

corresponding dual function value:

$$\begin{aligned} g(\lambda^*(t)) &= f_0(x^*(t)) - \sum_{i=1}^{L} \lambda^*(t)_i^T f_i(x^*(t)) \\ &= f_0(x^*(t)) - m/t \end{aligned}$$

where $m = \sum_{i=1}^{L} \theta_i$

# SUMT with generalized inequalities

**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 $t$

- **exactly the same** as standard SUMT
- with $m = \sum_i \theta_i$, everything generalizes, including
  - complexity analysis
  - typical performance



# Example: SDP

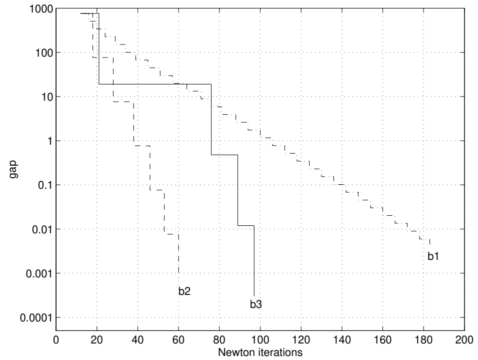
$$\begin{array}{ll} \text{minimize} & c^T x \\ \text{subject to} & F(x) = F_0 + x_1 F_1 + \cdots + x_n F_n \preceq 0 \end{array}$$

$x \in \mathbf{R}^n$, $F(x) \in \mathbf{R}^{p \times p}$

$$x^*(t) = \operatorname*{\mathbf{argmin}}_{F(z) \succ 0} \left( tc^T z - \log\det(-F(z)) \right)$$

self-concordant for all $t$ ($\theta = p$)

example with $n = 50$, $p = 100$
backtracking line search ($\alpha = 0.2$, $\beta = 0.5$)



## Summary

- other IP methods similar to SUMT
- work very well in practice
- worst-case complexity theory (if self-concordant)
- worst-case theory: # iters $O(\sqrt{m})$
- in practice: **much slower growth** than $\sqrt{m}$

sophisticated variations:

- use predictor steps to follow central path, with aggressive step size rules (*e.g.*, $99\%$ to boundary!)
- primal-dual methods
- infeasible methods (combine phase I & II)
- incomplete centering



## Summary of ideas

The two methods we have seen are typical of many interior-point methods

- they compute points on central path (outer iterations) using Newton's method (inner iterations)

  sometimes called 'path-following' methods
- outer iterations yield dual feasible points, lower bound on $f^\star$
- algorithm parameter controls tradeoff between $\#$ inner (Newton) steps required vs.$\#$ outer steps
- proof of convergence simple
- can work very well in practice (with minor modification, *e.g.*, $q \approx m$)

Problems/ unclear points:

- no clear way to understand tradeoff between # inner vs. # outer iterations
- not clear which barrier to use or why — ideas apply equally well to

$$\psi(x) = \begin{cases} \sum_i \dfrac{1}{-f_i(x)} & x \in C \\ +\infty & \text{otherwise} \end{cases}$$

  as logarithmic barrier $\phi$ ...

  (practice shows choice of barrier makes huge difference)
- no worst-case complexity analysis