Convex Optimization M2

Lecture 5

Barrier Method

Barrier Method

- inequality constrained minimization
- logarithmic barrier function and central path
- barrier method
- feasibility and phase I methods
- complexity analysis via self-concordance
- generalized inequalities

Inequality constrained minimization

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

- f_i convex, twice continuously differentiable
- $A \in \mathbb{R}^{p \times n}$ with $\mathbf{Rank}\, A = p$
- we assume $p^\star$ is finite and attained
- we assume problem is strictly feasible: there exists $\tilde{x}$ with

$$
\tilde{x} \in \mathbf{dom}\, f_0, \qquad f_i(\tilde{x}) < 0, \quad i = 1, \ldots, m, \qquad A\tilde{x} = b
$$

hence, strong duality holds and dual optimum is attained

Examples

- LP, QP, QCQP, GP
- entropy maximization with linear inequality constraints

$$\begin{array}{ll} \text{minimize} & \sum_{i=1}^{n} x_i \log x_i \\ \text{subject to} & Fx \preceq g \\ & Ax = b \end{array}$$

  with $\mathbf{dom}\, f_0 = \mathbb{R}^n_{++}$
- differentiability may require reformulating the problem, $e.g.$, piecewise-linear minimization or ℓ_∞-norm approximation via LP
- SDPs and SOCPs are better handled as problems with generalized inequalities (see later)

Logarithmic barrier

reformulation of (1) via indicator function:

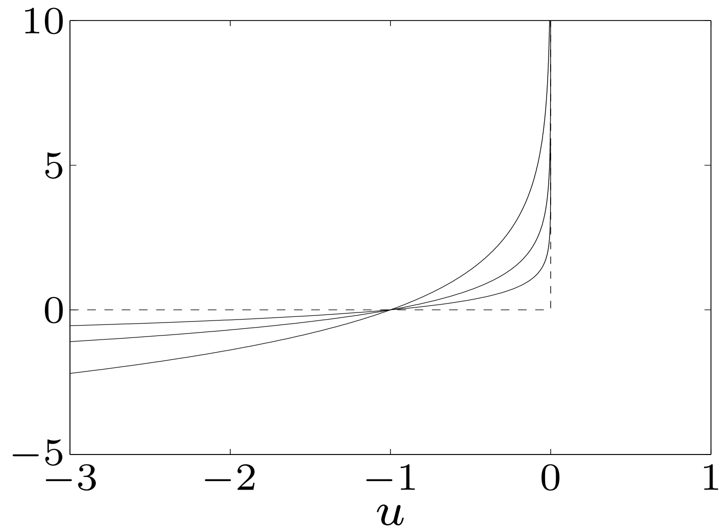
$$\begin{array}{ll} \text{minimize} & f_0(x) + \sum_{i=1}^m I_-(f_i(x)) \\ \text{subject to} & Ax = b \end{array}$$

where $I_-(u) = 0$ if $u \leq 0$, $I_-(u) = \infty$ otherwise (indicator function of $\mathbb{R}_-$)

approximation via logarithmic barrier

$$\begin{array}{ll} \text{minimize} & f_0(x) - (1/t)\sum_{i=1}^m \log(-f_i(x)) \\ \text{subject to} & Ax = b \end{array}$$

- an equality constrained problem
- for $t > 0$, $-(1/t)\log(-u)$ is a smooth approximation of I_-
- approximation improves as $t \to \infty$



logarithmic barrier function

$$\phi(x) = -\sum_{i=1}^{m} \log(-f_i(x)), \quad \mathbf{dom}\,\phi = \{x \mid f_1(x) < 0, \ldots, f_m(x) < 0\}$$

- convex (follows from composition rules)
- twice continuously differentiable, with derivatives

$$\begin{aligned}
\nabla\phi(x) &= \sum_{i=1}^{m} \frac{1}{-f_i(x)} \nabla f_i(x) \\
\nabla^2\phi(x) &= \sum_{i=1}^{m} \frac{1}{f_i(x)^2} \nabla f_i(x) \nabla f_i(x)^T + \sum_{i=1}^{m} \frac{1}{-f_i(x)} \nabla^2 f_i(x)
\end{aligned}$$

Central path

- for $t > 0$, define $x^\star(t)$ as the solution of

$$\begin{array}{ll} \text{minimize} & tf_0(x) + \phi(x) \\ \text{subject to} & Ax = b \end{array}$$

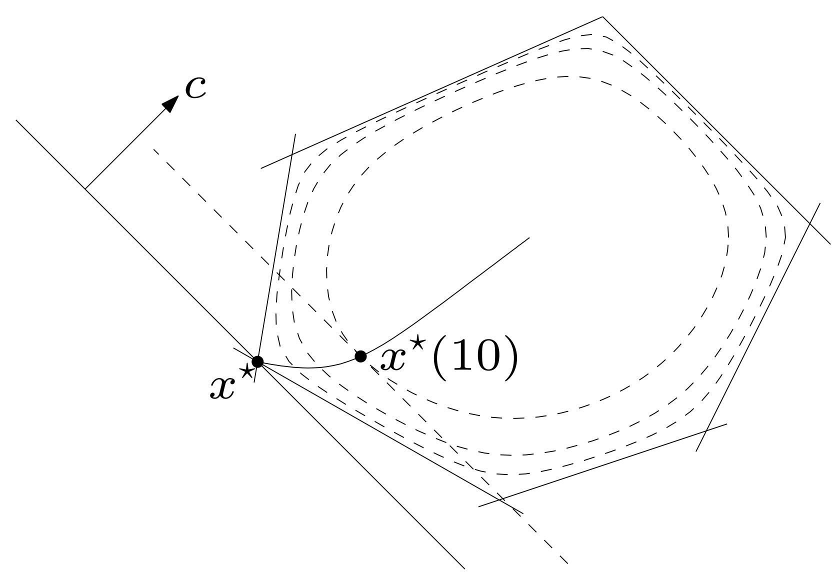
  (for now, assume $x^\star(t)$ exists and is unique for each $t > 0$)

- central path is $\{x^\star(t) \mid t > 0\}$

example: central path for an LP

$$\begin{array}{ll} \text{minimize} & c^T x \\ \text{subject to} & a_i^T x \leq b_i, \quad i = 1, \ldots, 6 \end{array}$$

hyperplane $c^T x = c^T x^\star(t)$ is tangent to level curve of ϕ through $x^\star(t)$



Dual points on central path

$x = x^\star(t)$ if there exists a w such that

$$t\nabla f_0(x) + \sum_{i=1}^m \frac{1}{-f_i(x)} \nabla f_i(x) + A^T w = 0, \qquad Ax = b$$

- therefore, $x^\star(t)$ minimizes the Lagrangian

$$L(x, \lambda^\star(t), \nu^\star(t)) = f_0(x) + \sum_{i=1}^m \lambda_i^\star(t) f_i(x) + \nu^\star(t)^T(Ax - b)$$

  where we define $\lambda_i^\star(t) = 1/(-tf_i(x^\star(t))$ and $\nu^\star(t) = w/t$
- this confirms the intuitive idea that $f_0(x^\star(t)) \to p^\star$ if $t \to \infty$:

$$\begin{array}{rcl} p^\star & \geq & g(\lambda^\star(t), \nu^\star(t)) \\ & = & L(x^\star(t), \lambda^\star(t), \nu^\star(t)) \\ & = & f_0(x^\star(t)) - m/t \end{array}$$

Interpretation via KKT conditions

$x = x^\star(t)$, $\lambda = \lambda^\star(t)$, $\nu = \nu^\star(t)$ satisfy

1. primal constraints: $f_i(x) \leq 0$, $i = 1, \ldots, m$, $Ax = b$
2. dual constraints: $\lambda \succeq 0$
3. approximate complementary slackness: $-\lambda_i f_i(x) = 1/t$, $i = 1, \ldots, m$
4. gradient of Lagrangian with respect to x vanishes:

$$\nabla f_0(x) + \sum_{i=1}^m \lambda_i \nabla f_i(x) + A^T \nu = 0$$

difference with KKT is that condition 3 replaces $\lambda_i f_i(x) = 0$

Force field interpretation

centering problem (for problem with no equality constraints)

$$\text{minimize} \quad t f_0(x) - \sum_{i=1}^m \log(-f_i(x))$$

force field interpretation

- $t f_0(x)$ is potential of force field $F_0(x) = -t\nabla f_0(x)$
- $-\log(-f_i(x))$ is potential of force field $F_i(x) = (1/f_i(x))\nabla f_i(x)$

the forces balance at $x^\star(t)$:

$$F_0(x^\star(t)) + \sum_{i=1}^m F_i(x^\star(t)) = 0$$

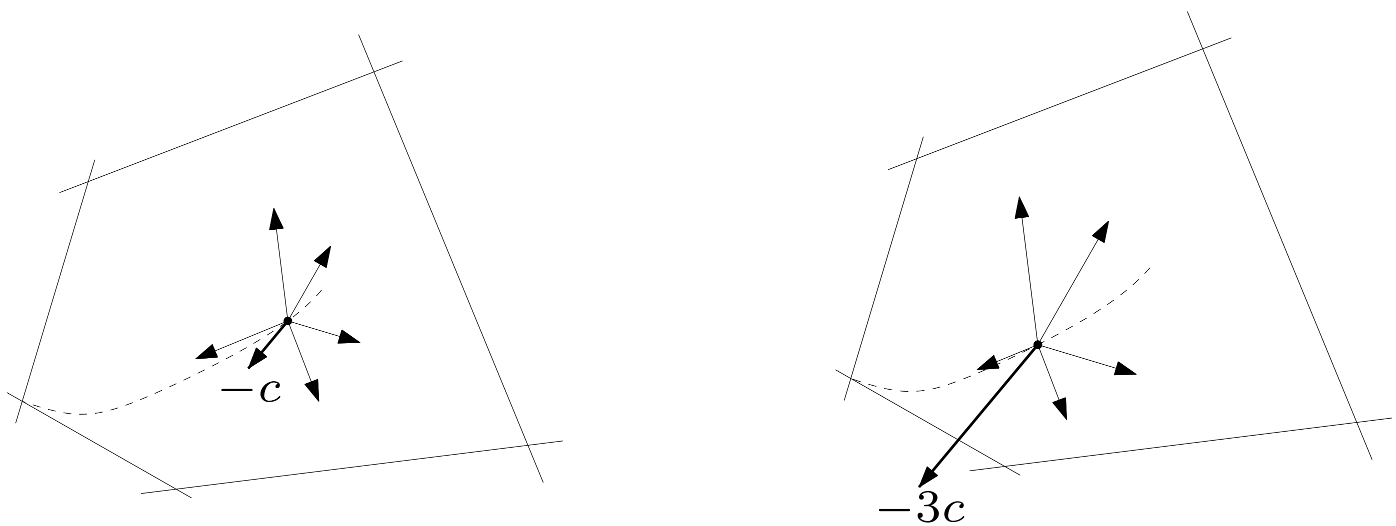
example

$$\begin{array}{ll} \text{minimize} & c^T x \\ \text{subject to} & a_i^T x \leq b_i, \quad i = 1, \ldots, m \end{array}$$

- objective force field is constant: $F_0(x) = -tc$
- constraint force field decays as inverse distance to constraint hyperplane:

$$F_i(x) = \frac{-a_i}{b_i - a_i^T x}, \qquad \|F_i(x)\|_2 = \frac{1}{\mathbf{dist}(x, \mathcal{H}_i)}$$

where $\mathcal{H}_i = \{x \mid a_i^T x = b_i\}$



$t = 1$ $\qquad\qquad$ $t = 3$

Barrier method

given strictly feasible x, $t := t^{(0)} > 0$, $\mu > 1$, tolerance $\epsilon > 0$.

repeat

1. *Centering step.* Compute $x^\star(t)$ by minimizing $tf_0 + \phi$, subject to $Ax = b$.
2. *Update.* $x := x^\star(t)$.
3. *Stopping criterion.* **quit** if $m/t < \epsilon$.
4. *Increase* t. $t := \mu t$.

- terminates with $f_0(x) - p^\star \leq \epsilon$ (stopping criterion follows from $f_0(x^\star(t)) - p^\star \leq m/t$)
- centering usually done using Newton's method, starting at current x
- choice of μ involves a trade-off: large μ means fewer outer iterations, more inner (Newton) iterations; typical values: $\mu = 10$–20
- several heuristics for choice of $t^{(0)}$

Convergence analysis

number of outer (centering) iterations: exactly

$$\left\lceil \frac{\log(m/(\epsilon t^{(0)}))}{\log \mu} \right\rceil$$

plus the initial centering step (to compute $x^\star(t^{(0)})$)

centering problem

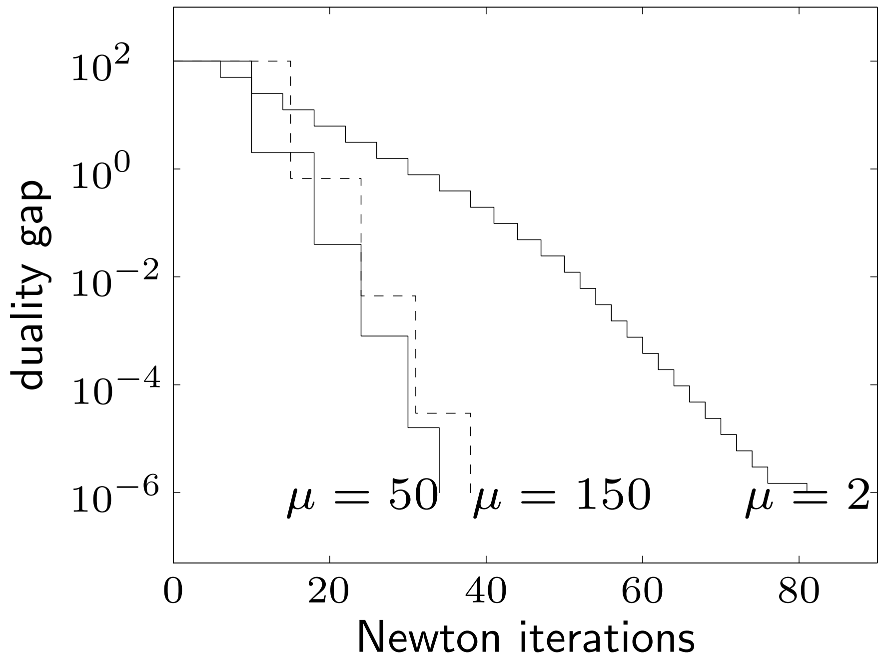
$$\text{minimize} \quad tf_0(x) + \phi(x)$$

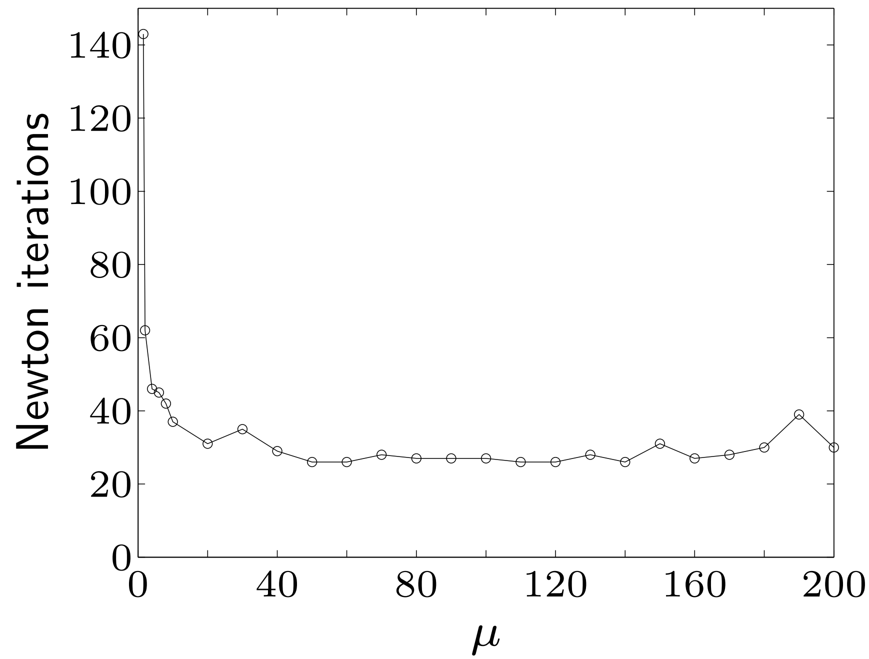
see convergence analysis of Newton's method

- $tf_0 + \phi$ must have closed sublevel sets for $t \geq t^{(0)}$
- classical analysis requires strong convexity, Lipschitz condition
- analysis via self-concordance requires self-concordance of $tf_0 + \phi$

Examples

inequality form LP ($m = 100$ inequalities, $n = 50$ variables)

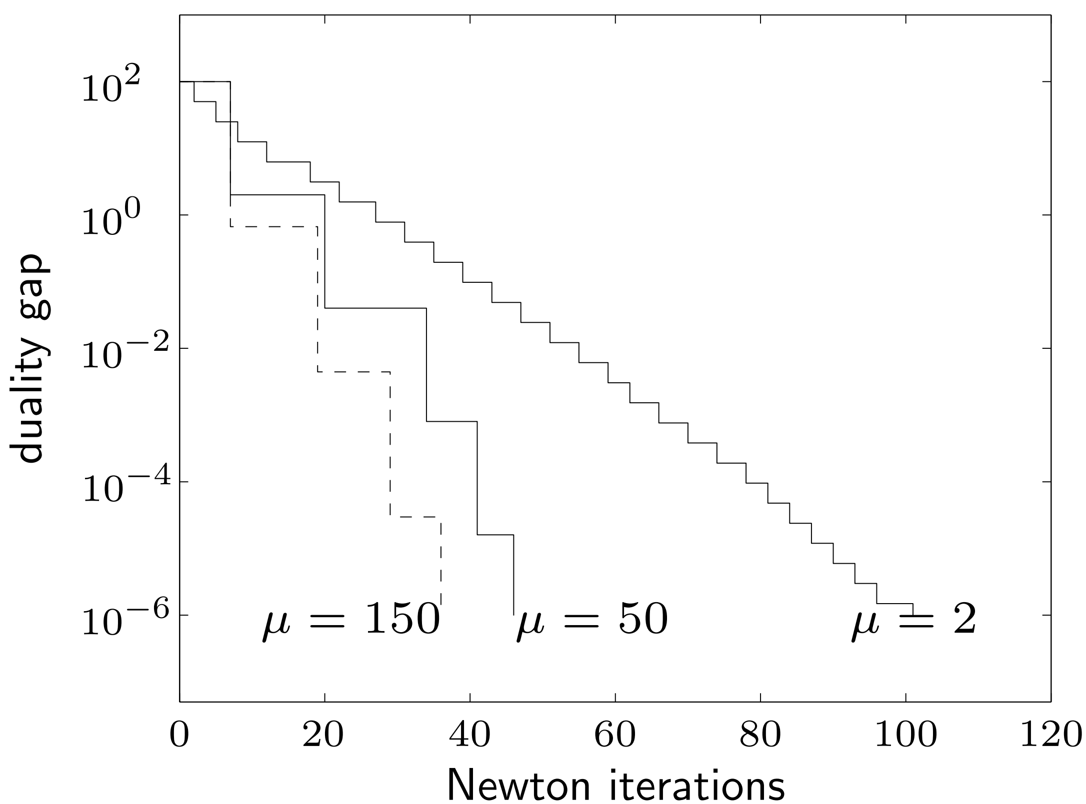




- starts with x on central path ($t^{(0)} = 1$, duality gap 100)
- terminates when $t = 10^8$ (gap 10^{-6})
- centering uses Newton's method with backtracking
- total number of Newton iterations not very sensitive for $\mu \geq 10$

geometric program ($m = 100$ inequalities and $n = 50$ variables)

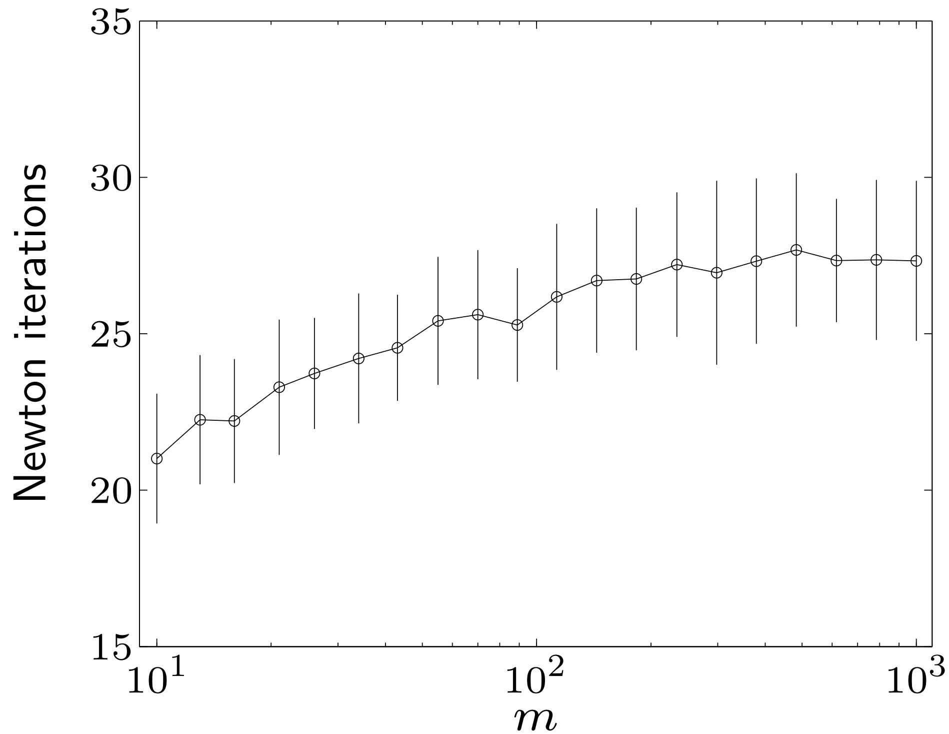
$$\begin{array}{ll} \text{minimize} & \log\left(\sum_{k=1}^{5} \exp(a_{0k}^T x + b_{0k})\right) \\ \text{subject to} & \log\left(\sum_{k=1}^{5} \exp(a_{ik}^T x + b_{ik})\right) \leq 0, \quad i = 1, \ldots, m \end{array}$$



family of standard LPs ($A \in \mathbb{R}^{m \times 2m}$)

$$\begin{array}{ll} \text{minimize} & c^T x \\ \text{subject to} & Ax = b, \quad x \succeq 0 \end{array}$$

$m = 10, \ldots, 1000$; for each m, solve 100 randomly generated instances



number of iterations grows very slowly as m ranges over a $100 : 1$ ratio

Feasibility and phase I methods

feasibility problem: find x such that

$$f_i(x) \le 0, \quad i = 1, \dots, m, \qquad Ax = b \tag{2}$$

phase I: computes strictly feasible starting point for barrier method

basic phase I method

$$\begin{array}{ll} \text{minimize (over } x,\ s) & s \\ \text{subject to} & f_i(x) \le s, \quad i = 1, \dots, m \\ & Ax = b \end{array} \tag{3}$$

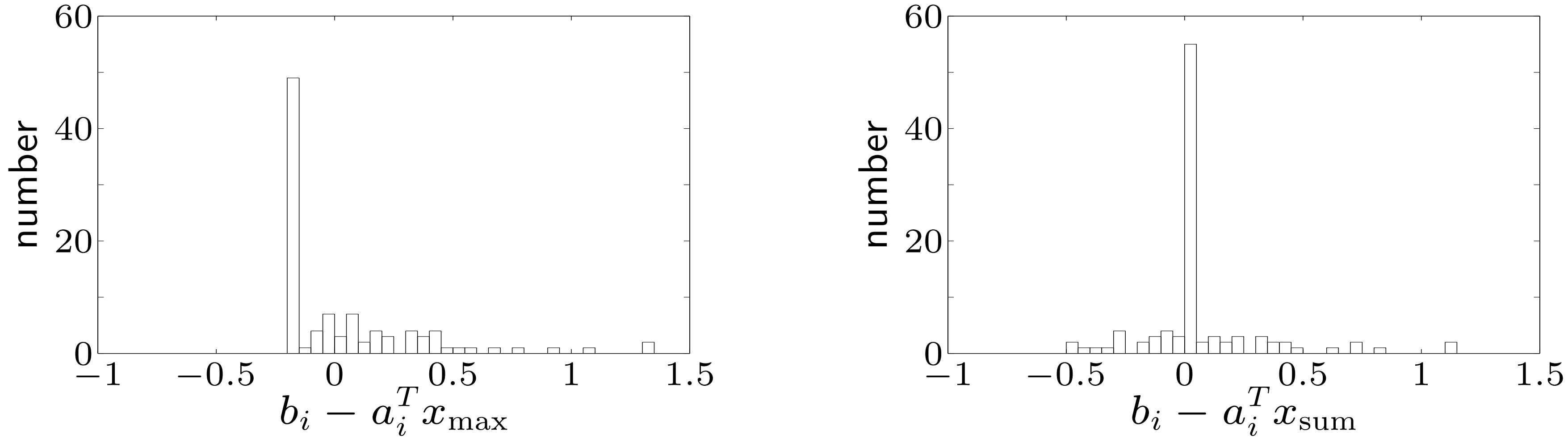
- if x, s feasible, with $s < 0$, then x is strictly feasible for (2)
- if optimal value $\bar{p}^\star$ of (3) is positive, then problem (2) is infeasible
- if $\bar{p}^\star = 0$ and attained, then problem (2) is feasible (but not strictly); if $\bar{p}^\star = 0$ and not attained, then problem (2) is infeasible

sum of infeasibilities phase I method

$$
\begin{array}{ll}
\text{minimize} & \mathbf{1}^T s \\
\text{subject to} & s \succeq 0, \quad f_i(x) \leq s_i, \quad i = 1, \ldots, m \\
& Ax = b
\end{array}
$$

for infeasible problems, produces a solution that satisfies many more inequalities than basic phase I method

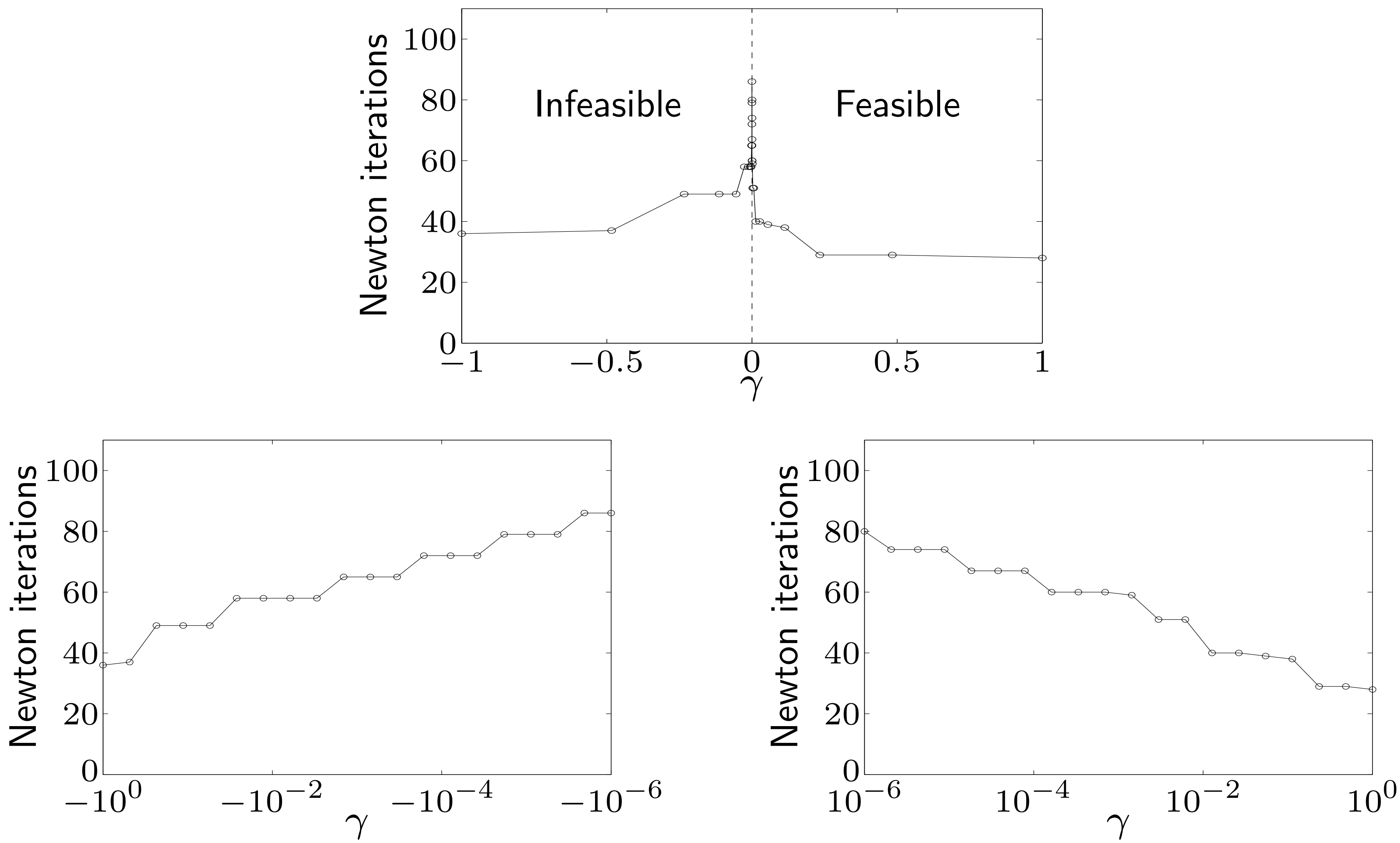
example (infeasible set of 100 linear inequalities in 50 variables)



left: basic phase I solution; satisfies 39 inequalities
right: sum of infeasibilities phase I solution; satisfies 79 inequalities

example: family of linear inequalities $Ax \preceq b + \gamma \Delta b$

- data chosen to be strictly feasible for $\gamma > 0$, infeasible for $\gamma \leq 0$
- use basic phase I, terminate when $s < 0$ or dual objective is positive



number of iterations roughly proportional to $\log(1/|\gamma|)$

Complexity analysis via self-concordance

same assumptions as on page 4, plus:

- sublevel sets (of f_0, on the feasible set) are bounded
- $tf_0 + \phi$ is self-concordant with closed sublevel sets

second condition

- holds for LP, QP, QCQP
- may require reformulating the problem, $e.g.$,

$$\begin{array}{ll} \text{minimize} & \sum_{i=1}^n x_i \log x_i \\ \text{subject to} & Fx \preceq g \end{array} \quad \longrightarrow \quad \begin{array}{ll} \text{minimize} & \sum_{i=1}^n x_i \log x_i \\ \text{subject to} & Fx \preceq g, \quad x \succeq 0 \end{array}$$

- needed for complexity analysis; barrier method works even when self-concordance assumption does not apply

Newton iterations per centering step: from self-concordance theory

$$\#\text{Newton iterations} \leq \frac{\mu t f_0(x) + \phi(x) - \mu t f_0(x^+) - \phi(x^+)}{\gamma} + c$$

- bound on effort of computing $x^+ = x^\star(\mu t)$ starting at $x = x^\star(t)$
- γ, c are constants (depend only on Newton algorithm parameters)
- from duality (with $\lambda = \lambda^\star(t)$, $\nu = \nu^\star(t)$):

$$\begin{array}{rl} & \mu t f_0(x) + \phi(x) - \mu t f_0(x^+) - \phi(x^+) \\ = & \mu t f_0(x) - \mu t f_0(x^+) + \displaystyle\sum_{i=1}^m \log(-\mu t \lambda_i f_i(x^+)) - m \log \mu \\ \leq & \mu t f_0(x) - \mu t f_0(x^+) - \mu t \displaystyle\sum_{i=1}^m \lambda_i f_i(x^+) - m - m \log \mu \\ \leq & \mu t f_0(x) - \mu t g(\lambda, \nu) - m - m \log \mu \\ = & m(\mu - 1 - \log \mu) \end{array}$$

total number of Newton iterations (excluding first centering step)

$$\#\text{Newton iterations} \leq N = \left\lceil \frac{\log(m/(t^{(0)}\epsilon))}{\log \mu} \right\rceil \left(\frac{m(\mu - 1 - \log \mu)}{\gamma} + c \right)$$

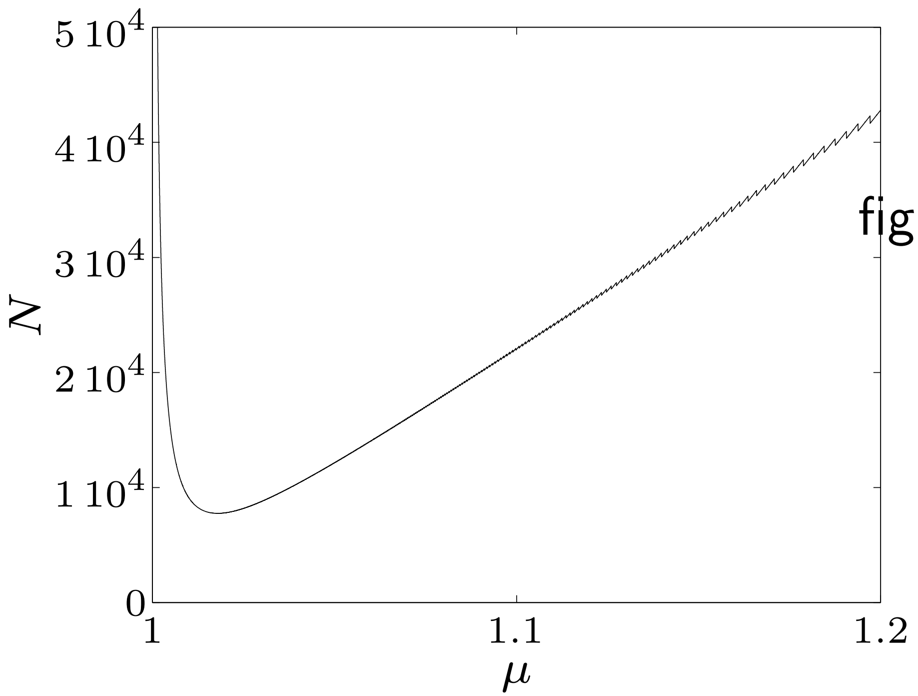


figure shows N for typical values of γ, c,

$$m = 100, \qquad \frac{m}{t^{(0)}\epsilon} = 10^5$$

- confirms trade-off in choice of μ
- in practice, #iterations is in the tens; not very sensitive for $\mu \geq 10$

polynomial-time complexity of barrier method

- for $\mu = 1 + 1/\sqrt{m}$:

$$N = O\left(\sqrt{m}\log\left(\frac{m/t^{(0)}}{\epsilon}\right)\right)$$

- number of Newton iterations for fixed gap reduction is $O(\sqrt{m})$
- multiply with cost of one Newton iteration (a polynomial function of problem dimensions), to get bound on number of flops

this choice of μ optimizes worst-case complexity; in practice we choose μ fixed ($\mu = 10, \ldots, 20$)

Generalized inequalities

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

- f_0 convex, $f_i : \mathbb{R}^n \to \mathbb{R}^{k_i}$, $i=1,\ldots,m$, convex with respect to proper cones $K_i \in \mathbb{R}^{k_i}$
- f_i twice continuously differentiable
- $A \in \mathbb{R}^{p\times n}$ with $\mathbf{Rank}\, A = p$
- we assume $p^\star$ is finite and attained
- we assume problem is strictly feasible; hence strong duality holds and dual optimum is attained

examples of greatest interest: SOCP, SDP

Generalized logarithm for proper cone

$\psi : \mathbb{R}^q \to \mathbb{R}$ is generalized logarithm for proper cone $K \subseteq \mathbb{R}^q$ if:

- $\mathbf{dom}\,\psi = \mathbf{int}\,K$ and $\nabla^2\psi(y) \prec 0$ for $y \succ_K 0$
- $\psi(sy) = \psi(y) + \theta \log s$ for $y \succ_K 0$, $s > 0$ (θ is the degree of ψ)

examples

- nonnegative orthant $K = \mathbb{R}^n_+$: $\psi(y) = \sum_{i=1}^n \log y_i$, with degree $\theta = n$
- positive semidefinite cone $K = \mathbf{S}^n_+$:

$$\psi(Y) = \log\det Y \qquad (\theta = n)$$

- second-order cone $K = \{y \in \mathbb{R}^{n+1} \mid (y_1^2 + \cdots + y_n^2)^{1/2} \leq y_{n+1}\}$:

$$\psi(y) = \log(y_{n+1}^2 - y_1^2 - \cdots - y_n^2) \qquad (\theta = 2)$$

properties (without proof): for $y \succ_K 0$,

$$\nabla\psi(y) \succeq_{K^*} 0, \qquad y^T\nabla\psi(y) = \theta$$

- nonnegative orthant $\mathbb{R}^n_+$: $\psi(y) = \sum_{i=1}^n \log y_i$

$$\nabla\psi(y) = (1/y_1, \ldots, 1/y_n), \qquad y^T\nabla\psi(y) = n$$

- positive semidefinite cone $\mathbf{S}^n_+$: $\psi(Y) = \log\det Y$

$$\nabla\psi(Y) = Y^{-1}, \qquad \mathbf{Tr}(Y\nabla\psi(Y)) = n$$

- second-order cone $K = \{y \in \mathbb{R}^{n+1} \mid (y_1^2 + \cdots + y_n^2)^{1/2} \leq y_{n+1}\}$:

$$\psi(y) = \frac{2}{y_{n+1}^2 - y_1^2 - \cdots - y_n^2}\begin{bmatrix} -y_1 \\ \vdots \\ -y_n \\ y_{n+1} \end{bmatrix}, \qquad y^T\nabla\psi(y) = 2$$

Logarithmic barrier and central path

logarithmic barrier for $f_1(x) \preceq_{K_1} 0, \ldots, f_m(x) \preceq_{K_m} 0$:

$$\phi(x) = -\sum_{i=1}^{m} \psi_i(-f_i(x)), \qquad \mathbf{dom}\,\phi = \{x \mid f_i(x) \prec_{K_i} 0,\ i = 1, \ldots, m\}$$

- ψ_i is generalized logarithm for K_i, with degree θ_i
- ϕ is convex, twice continuously differentiable

central path: $\{x^\star(t) \mid t > 0\}$ where $x^\star(t)$ solves

$$\begin{array}{ll} \text{minimize} & t f_0(x) + \phi(x) \\ \text{subject to} & Ax = b \end{array}$$

Dual points on central path

$x = x^\star(t)$ if there exists $w \in \mathbb{R}^p$,

$$t \nabla f_0(x) + \sum_{i=1}^m Df_i(x)^T \nabla \psi_i(-f_i(x)) + A^T w = 0$$

($Df_i(x) \in \mathbb{R}^{k_i \times n}$ is derivative matrix of f_i)

- therefore, $x^\star(t)$ minimizes Lagrangian $L(x, \lambda^\star(t), \nu^\star(t))$, where

$$\lambda_i^\star(t) = \frac{1}{t} \nabla \psi_i(-f_i(x^\star(t))), \qquad \nu^\star(t) = \frac{w}{t}$$

- from properties of ψ_i: $\lambda_i^\star(t) \succ_{K_i^*} 0$, with duality gap

$$f_0(x^\star(t)) - g(\lambda^\star(t), \nu^\star(t)) = (1/t) \sum_{i=1}^m \theta_i$$

example: semidefinite programming (with $F_i \in \mathbf{S}^p$)

$$\begin{array}{ll} \text{minimize} & c^T x \\ \text{subject to} & F(x) = \sum_{i=1}^n x_i F_i + G \preceq 0 \end{array}$$

- logarithmic barrier: $\phi(x) = \log\det(-F(x)^{-1})$
- central path: $x^\star(t)$ minimizes $tc^T x - \log\det(-F(x))$; hence

$$tc_i - \mathbf{Tr}(F_i F(x^\star(t))^{-1}) = 0, \quad i = 1, \ldots, n$$

- dual point on central path: $Z^\star(t) = -(1/t)F(x^\star(t))^{-1}$ is feasible for

$$\begin{array}{ll} \text{maximize} & \mathbf{Tr}(GZ) \\ \text{subject to} & \mathbf{Tr}(F_i Z) + c_i = 0, \quad i = 1, \ldots, n \\ & Z \succeq 0 \end{array}$$

- duality gap on central path: $c^T x^\star(t) - \mathbf{Tr}(GZ^\star(t)) = p/t$

Barrier method

given strictly feasible x, $t := t^{(0)} > 0$, $\mu > 1$, tolerance $\epsilon > 0$.

repeat

1. *Centering step.* Compute $x^\star(t)$ by minimizing $tf_0 + \phi$, subject to $Ax = b$.
2. *Update.* $x := x^\star(t)$.
3. *Stopping criterion.* **quit** if $(\sum_i \theta_i)/t < \epsilon$.
4. *Increase* t. $t := \mu t$.

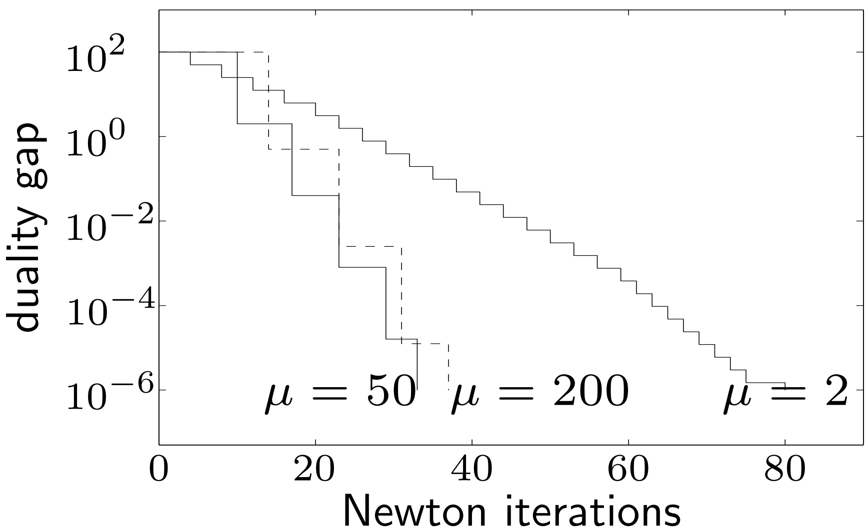
- only difference is duality gap m/t on central path is replaced by $\sum_i \theta_i/t$
- number of outer iterations:

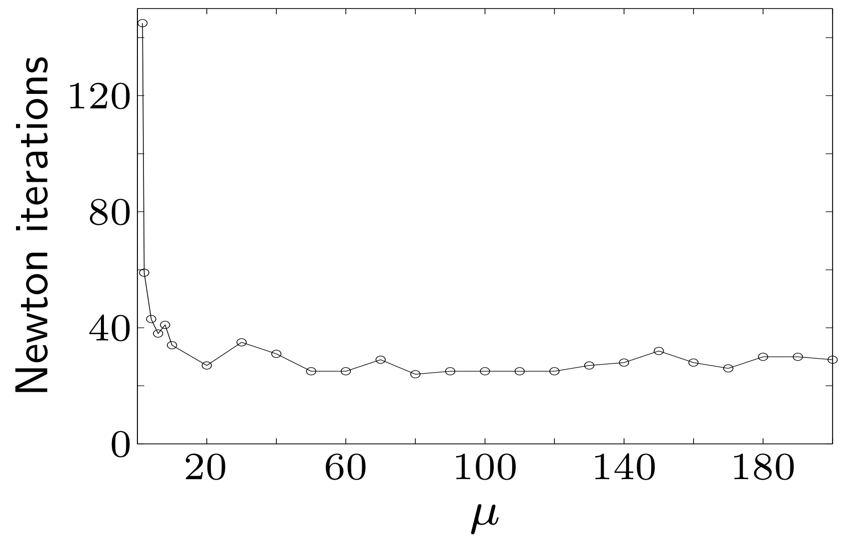
$$\left\lceil \frac{\log((\sum_i \theta_i)/(\epsilon t^{(0)}))}{\log \mu} \right\rceil$$

- complexity analysis via self-concordance applies to SDP, SOCP

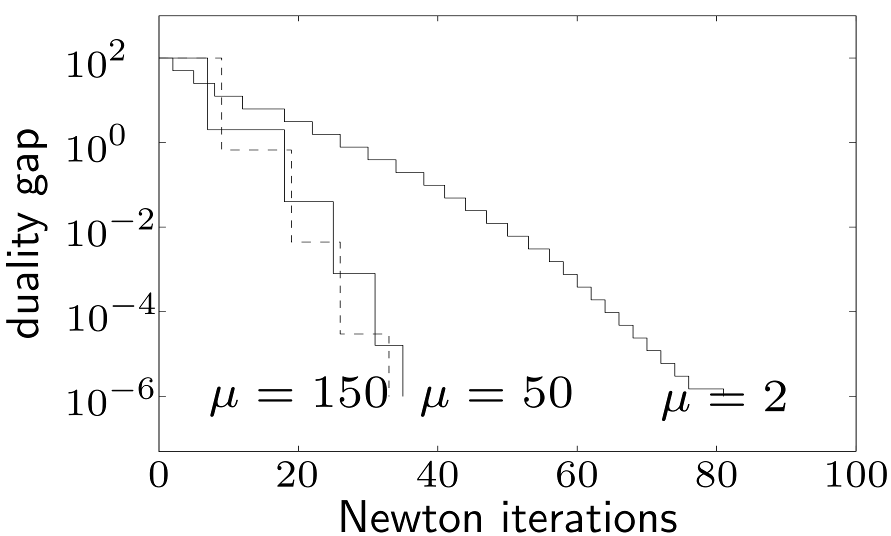
Examples

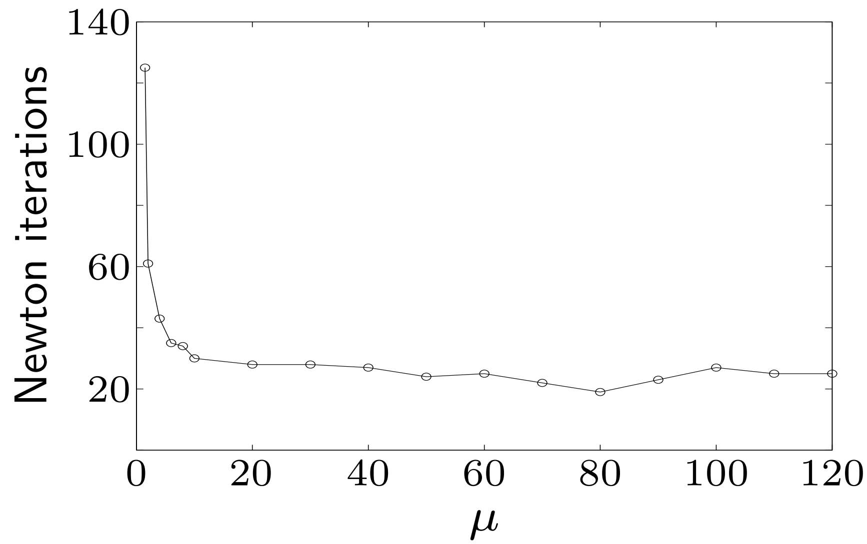
second-order cone program (50 variables, 50 SOC constraints in $\mathbb{R}^6$)





semidefinite program (100 variables, LMI constraint in $\mathbf{S}^{100}$)

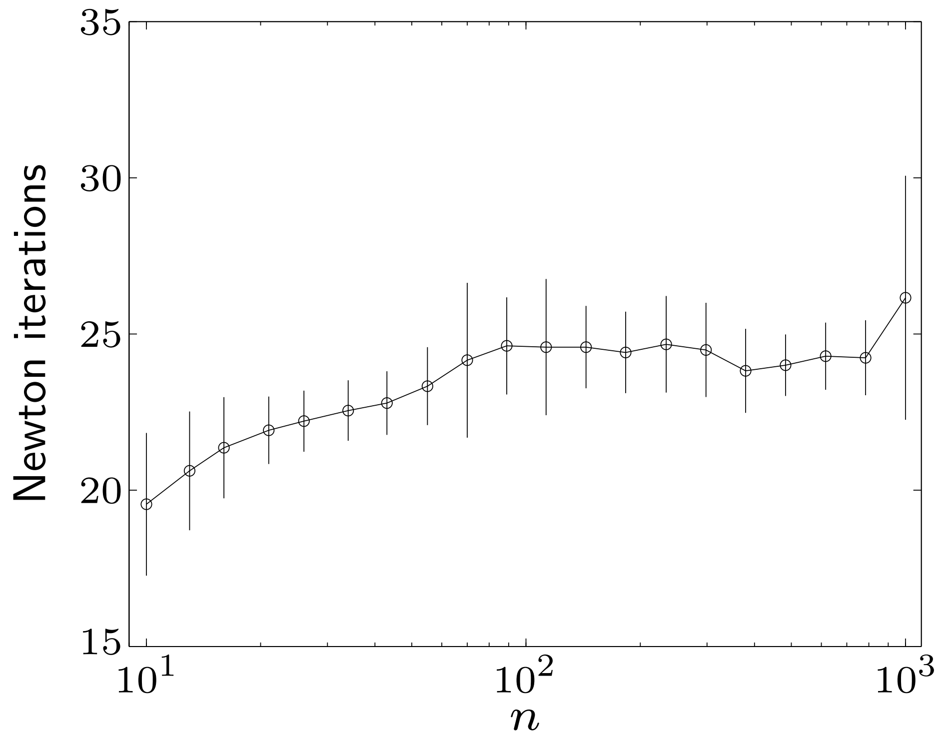




family of SDPs ($A \in \mathbf{S}^n$, $x \in \mathbb{R}^n$)

$$\begin{array}{ll} \text{minimize} & \mathbf{1}^T x \\ \text{subject to} & A + \mathbf{diag}(x) \succeq 0 \end{array}$$

$n = 10, \ldots, 1000$, for each n solve 100 randomly generated instances



Primal-dual interior-point methods

more efficient than barrier method when high accuracy is needed

- update primal and dual variables at each iteration; no distinction between inner and outer iterations
- often exhibit superlinear asymptotic convergence
- search directions can be interpreted as Newton directions for modified KKT conditions
- can start at infeasible points
- cost per iteration same as barrier method

Interior-point methods: summary

- Interior point methods (IPM) are very reliable on small scale problems.
  - Example: SDP of dimension 100, SOCP with less than a thousand variables.
  - Most conic problems with a couple of hundred variables can formulated and solved very quickly using preprocessors such as CVX.
- IPM often efficient on larger problems if KKT system has some structure (sparsity, blocks, etc).
  - Large scale linear programs with thousands of variables are routinely solved by free or commercial solvers using IPM (e.g. SDPT3, MOSEK, GLPK, CPLEX, etc.).
  - Much larger sparse LPs can also be solved efficiently using the same techniques.
- Not workable for very large problems.
  - For some problems, e.g. semidefinite programs, exploiting structure in IPM is hard.
  - First order methods (using the gradient only) seem to be the only option for extremely large problems

Semidefinite programming: CVX

Solving the maxcut relaxation

$$\begin{array}{ll}\text{max.} & \mathbf{Tr}(XC) \\ \text{s.t.} & \mathbf{diag}(X) = \mathbf{1} \\ & X \succeq 0,\end{array}$$

is written as follows in CVX/MATLAB

```
cvx_begin
.   variable X(n,n) symmetric
.   maximize trace(C*X)
.   subject to
.     diag(X)==1
.     X==semidefinite(n)
cvx_end
```