# Convex Optimization

Stephen Boyd    Lieven Vandenberghe

Revised slides by Stephen Boyd, Lieven Vandenberghe, and Parth Nobel

# 11. Interior-point methods

## Outline

## Inequality constrained minimization

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

we assume

- $f_i$ convex, twice continuously differentiable
- $A \in \mathbf{R}^{p \times n}$ with $\mathbf{rank}\, A = p$
- $p^\star$ is finite and attained
- 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

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

with $\mathbf{dom}\, f_0 = \mathbf{R}_{++}^n$

▶ differentiability may require reformulating the problem, *e.g.*, piecewise-linear minimization or $\ell_\infty$-norm approximation via LP

▶ SDPs and SOCPs are better handled as problems with generalized inequalities (see later)

## Outline

## Logarithmic barrier

- reformulation via **indicator function**:

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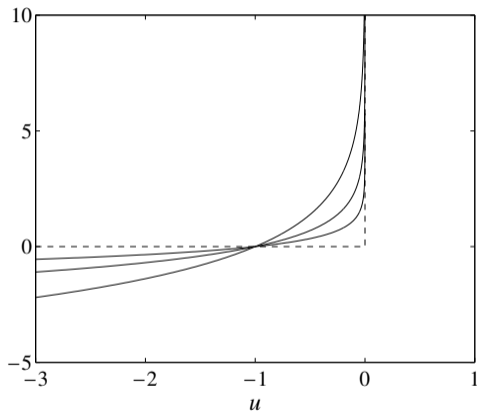\text{subject to} & Ax = b
\end{array}$$

where $I_-(u) = 0$ if $u \leq 0$, $I_-(u) = \infty$ otherwise

- **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 $t \to \infty$

- $-(1/t) \log u$ for three values of $t$, and $I_-(u)$

## Logarithmic barrier function

- ▶ log barrier function for constraints $f_1(x) \le 0, \ldots, f_m(x) \le 0$

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

- ▶ convex (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

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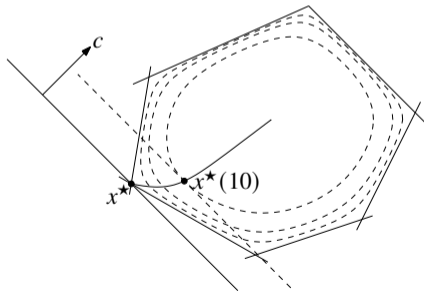\text{subject to} & Ax = b
\end{array}$$

(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 $\phi$ 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 $\lambda_i^\star(t) = 1/(-t f_i(x^\star(t)))$ and $\nu^\star(t) = w/t$

▶ this confirms the intuitive idea that $f_0(x^\star(t)) \to p^\star$ if $t \to \infty$:

$$p^\star \geq g(\lambda^\star(t), \nu^\star(t)) = L(x^\star(t), \lambda^\star(t), \nu^\star(t)) = f_0(x^\star(t)) - m/t$$

## Interpretation via KKT conditions

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

1. primal constraints: $f_i(x) \le 0$, $i = 1, \ldots, m$, $Ax = b$
2. dual constraints: $\lambda \ge 0$
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)

$$\text{minimize} \quad tf_0(x) - \sum_{i=1}^{m} \log(-f_i(x))$$

- **force field interpretation**
  - $tf_0(x)$ is potential of force field $F_0(x) = -t\nabla f_0(x)$

  - $-\log(-f_i(x))$ is potential of force field $F_i(x) = (1/f_i(x))\nabla f_i(x)$
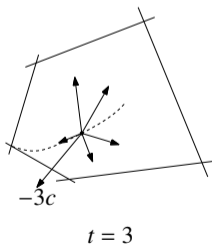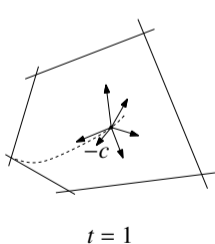
- forces balance at $x^\star(t)$:

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

## Example: LP

▶ minimize $c^T x$ subject to $a_i^T x \le b_i$, $i = 1, \ldots, m$, with $x \in \mathbf{R}^n$

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



$t = 1$        $t = 3$

## Outline

## 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 $\mu$ involves a trade-off: large $\mu$ means fewer outer iterations, more inner (Newton) iterations; typical values: $\mu = 10$ or $20$

▶ several heuristics for choice of $t^{(0)}$

## Example: Inequality form LP
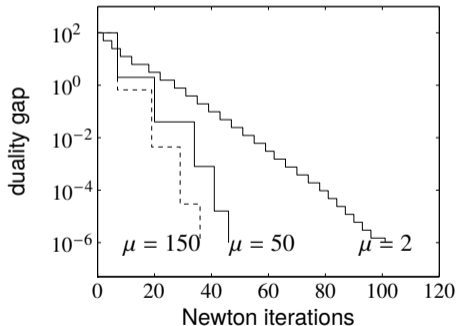
($m = 100$ inequalities, $n = 50$ variables)
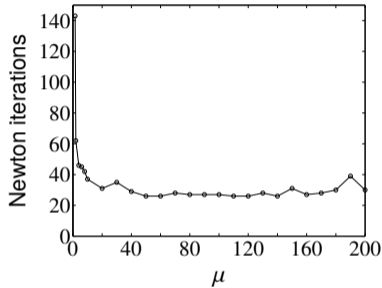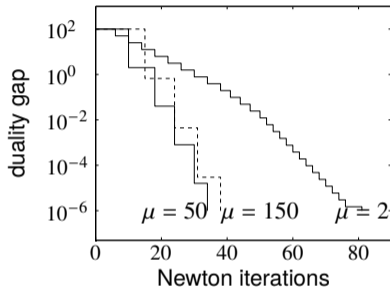


- starts with $x$ on central path ($t^{(0)} = 1$, duality gap 100)
- terminates when $t = 10^8$ (gap $10^{-6}$)
- total number of Newton iterations not very sensitive for $\mu \geq 10$

## Example: Geometric program in convex form

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

$$\text{minimize} \quad \log\left(\sum_{k=1}^{5} \exp(a_{0k}^T x + b_{0k})\right)$$
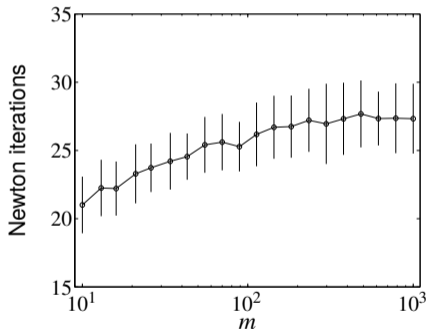$$\text{subject to} \quad \log\left(\sum_{k=1}^{5} \exp(a_{ik}^T x + b_{ik})\right) \le 0, \quad i = 1, \dots, m$$

## Family of standard LPs

($A \in \mathbf{R}^{m \times 2m}$)

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

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



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

## Outline

## Phase I methods

▶ barrier method needs strictly feasible starting point, *i.e.*, $x$ with

$$f_i(x) < 0, \quad i = 1, \ldots, m, \qquad Ax = b$$

▶ (like the infeasible start Newton method, more sophisticated interior-point methods do not require a feasible starting point)

▶ **phase I** method forms an optimization problem that
  – is itself strictly feasible
  – finds a strictly feasible point for original problem, if one exists
  – certifies original problem as infeasible otherwise

▶ **phase II** uses barrier method starting from strictly feasible point found in phase I

## Basic phase I method

▶ introduce slack variable $s$ in **phase I problem**

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

with optimal value $\bar{p}^\star$

– if $\bar{p}^\star < 0$, original inequalities are strictly feasible
– if $\bar{p}^\star > 0$, original inequalities are infeasible
– $\bar{p}^\star = 0$ is an ambiguous case

▶ start phase I problem with

– any $\tilde{x}$ in problem domain with $A\tilde{x} = b$
– $s = 1 + \max_i f_i(\tilde{x})$

# Sum of infeasibilities phase I method

▶ minimize **sum** of slacks, not max:

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

▶ will find a strictly feasible point if one exists

▶ for infeasible problems, produces a solution that satisfies many (but not all) inequalities

▶ can weight slacks to set **priorities** (in satifying constraints)

# 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
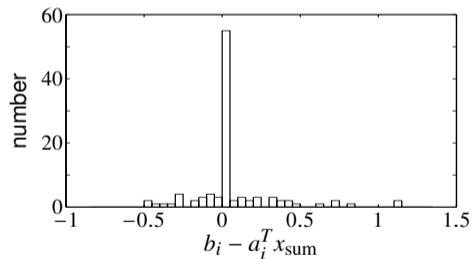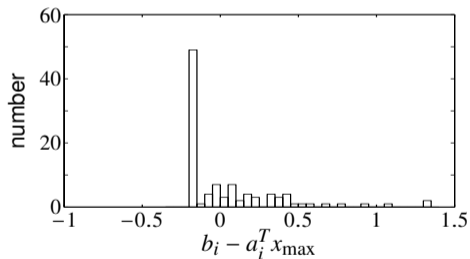
## Example: Family of linear inequalities

- $Ax \preceq b + \gamma \Delta b$; strictly feasible for $\gamma > 0$, infeasible for $\gamma < 0$
- use basic phase I, terminate when $s < 0$ or dual objective is positive
- number of iterations roughly proportional to $\log(1/|\gamma|)$
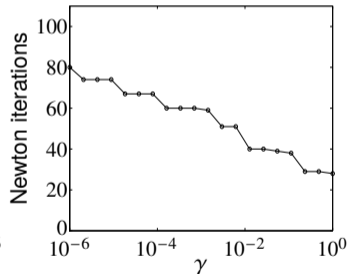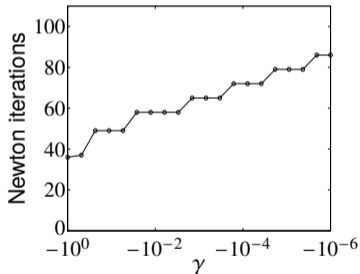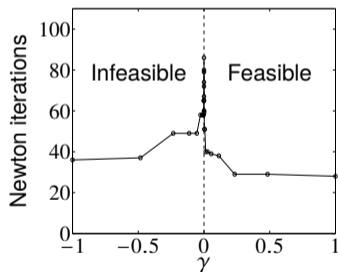
# Outline

# Number of outer iterations

▶ in each iteration duality gap is reduced by exactly the factor $\mu$

▶ **number of outer (centering) iterations** is exactly

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

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

▶ we will bound **number of Newton steps per centering iteration** using self-concordance analysis

## Complexity analysis via self-concordance

same assumptions as on slide 11.2, 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.*,

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

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

## Newton iterations per centering step

▶ we compute $x^+ = x^\star(\mu t)$, by minimizing $\mu t f_0(x) + \phi(x)$ starting from $x = x^\star(t)$

▶ from self-concordance theory,

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

▶ $\gamma, c$ are constants (that depend only on Newton algorithm parameters)

▶ we will bound numerator $\mu t f_0(x) + \phi(x) - \mu t f_0(x^+) - \phi(x^+)$

▶ with $\lambda_i = \lambda_i^\star(t) = -1/(t f_i(x))$, we have $-f_i(x) = 1/(t\lambda_i)$, so

$$\phi(x) = \sum_{i=1}^{m} -\log(-f_i(x)) = \sum_{i=1}^{m} \log(t\lambda_i)$$

so

$$\phi(x) - \phi(x^+) = \sum_{i=1}^{m} \left(\log(t\lambda_i) + \log(-f_i(x^+))\right) = \sum_{i=1}^{m} \log(-\mu t \lambda_i f_i(x^+)) - m \log \mu$$
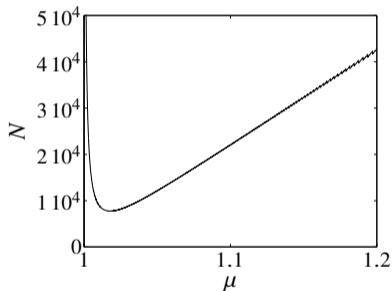
using $\log u \le u - 1$ we have $\phi(x) - \phi(x^+) \le -\mu t \sum_{i=1}^{m} \lambda_i f_i(x^+) - m - m \log \mu$, so

$$
\begin{aligned}
\mu t f_0(x) &+ \phi(x) - \mu t f_0(x^+) - \phi(x^+) \\
&\le \mu t f_0(x) - \mu t f_0(x^+) - \mu t \sum_{i=1}^{m} \lambda_i f_i(x^+) - m - m \log \mu \\
&= \mu t f_0(x) - \mu t \left( f_0(x^+) + \sum_{i=1}^{m} \lambda_i f_i(x^+) + \nu^T(Ax^+ - b) \right) - m - m \log \mu \\
&= \mu t f_0(x) - \mu t L(x^+, \lambda, \nu) - m - m \log \mu \\
&\le \mu t f_0(x) - \mu t g(\lambda, \nu) - m - m \log \mu \\
&= m(\mu - 1 - \log \mu)
\end{aligned}
$$

using $L(x^+, \lambda, nu) \ge g(\lambda, \nu)$ in second last line and $f_0(x) - g(\lambda, \nu) = m/t$ in last line

## Total number of Newton iterations

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



$N$ versus $\mu$ for typical values of $\gamma$, $c$;
$m = 100$, initial duality gap $\frac{m}{t^{(0)}\epsilon} = 10^5$

▶ confirms trade-off in choice of $\mu$

▶ 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 $\mu$ optimizes worst-case complexity; in practice we choose $\mu$ fixed and larger

## **Outline**

## 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, $f_i : \mathbf{R}^n \to \mathbf{R}^{k_i}$, $i = 1, \ldots, m$, convex with respect to proper cones $K_i \in \mathbf{R}^{k_i}$

▶ we assume
  - $f_i$ twice continuously differentiable
  - $A \in \mathbf{R}^{p \times n}$ with $\mathbf{rank}\, A = p$
  - $p^\star$ is finite and attained
  - problem is strictly feasible; hence strong duality holds and dual optimum is attained

▶ examples of greatest interest: SOCP, SDP

# Generalized logarithm for proper cone

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

- **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 $y \succ_K 0$, $s > 0$ ($\theta$ is the degree of $\psi$)

**examples**

- nonnegative orthant $K = \mathbf{R}_+^n$: $\psi(y) = \sum_{i=1}^n \log y_i$, with degree $\theta = n$
- positive semidefinite cone $K = \mathbf{S}_+^n$: $\psi(Y) = \log \det Y$, with degree $\theta = n$
- second-order cone $K = \{y \in \mathbf{R}^{n+1} \mid (y_1^2 + \cdots + y_n^2)^{1/2} \le y_{n+1}\}$:

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

## Properties

▶ (without proof): for $y >_K 0$,

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

▶ nonnegative orthant $\mathbf{R}_+^n$: $\psi(y) = \sum_{i=1}^n \log y_i$

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

▶ 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 $K = \{y \in \mathbf{R}^{n+1} \mid (y_1^2 + \cdots + y_n^2)^{1/2} \le y_{n+1}\}$:

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

## Logarithmic barrier and central path

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

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

▶ $\psi_i$ is generalized logarithm for $K_i$, with degree $\theta_i$

▶ $\phi$ is convex, twice continuously differentiable

**central path:** $\{x^\star(t) \mid t > 0\}$ where $x^\star(t)$ is solution of

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

## Dual points on central path

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

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

($Df_i(x) \in \mathbf{R}^{k_i \times n}$ 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 $\psi_i$: $\lambda_i^\star(t) >_{K_i^*} 0$, 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: $c^T x^\star(t) - \mathbf{tr}(GZ^\star(t)) = p/t$

Boyd and Vandenberghe

## 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 $(\sum_i \theta_i)/t < \epsilon$.
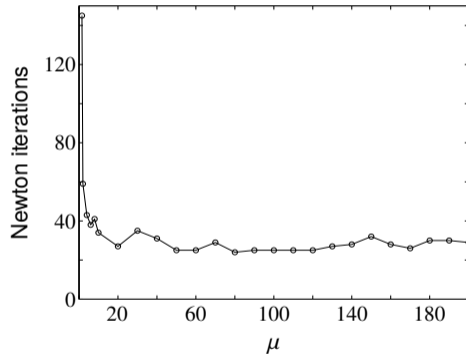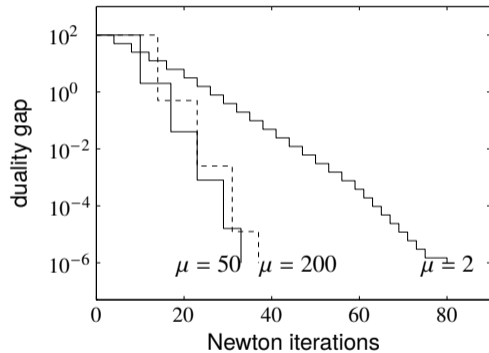4. *Increase $t$.* $t := \mu t$.

▶ only difference is duality gap $m/t$ on central path is replaced by $\sum_i \theta_i / t$

▶ number of outer iterations:

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

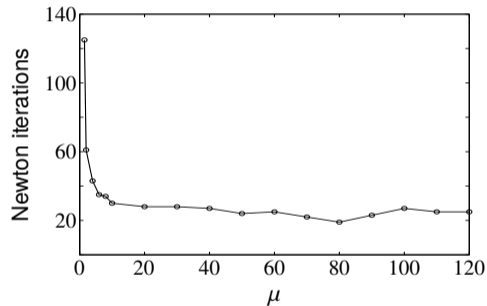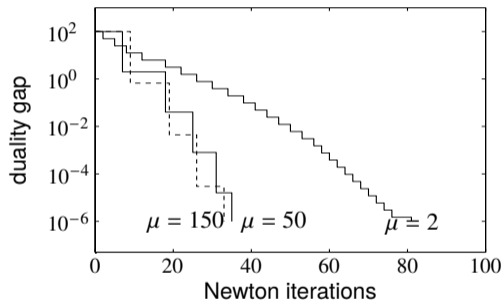▶ complexity analysis via self-concordance applies to SDP, SOCP

## Example: SOCP

(50 variables, 50 SOC constraints in $\mathbf{R}^6$)

## Example: SDP

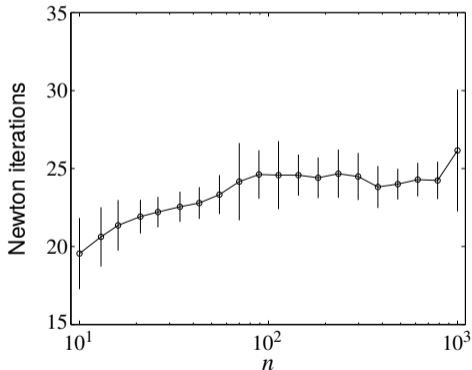(100 variables, LMI constraint in $\mathbf{S}^{100}$)

## Example: Family of SDPs

($A \in \mathbf{S}^n$, $x \in \mathbf{R}^n$)

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

$n = 10, \ldots, 1000$; 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, and $\kappa$, 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