Convex Optimization Theory and Applications

Topic 17 - Inequality Constrained Minimization

Li Li

Department of Automation
Tsinghua University

Fall, 2009-2021.

17.0. Outline

17.1. Inequality Constrained Minimization Problems

17.2. Barrier Method

17.3. Primal-Dual Interior-Point Methods

17.1. Ineq. Const. Minimization Problems

Assuming all problems are convex, you can think of the following hierarchy that we've worked through:

- Quadratic problems are the easiest: closed-form solution
- Equality-constrained quadratic problems are still easy: we use KKT conditions to derive closed-form solution
- Equality-constrained smooth problems are next: use Newton's method to reduce this to a sequence of equality-constrained quadratic problems
- Inequality-constrained and also equality-constrained smooth problems are what we cover now: use interior-point methods to reduce this to a sequence of equality-constrained problems

17.1. Ineq. Const. Minimization Problems

含不等式约束（凸）优化问题

$$\min\left\{ f_0(x) \middle| \text{ s.t. } f_i(x) \le 0,\ i = 1, \ldots, m,\ Ax = b \right\}$$

$f_i : R^n \mapsto R$， 定义域为 $\mathrm{dom}\ f_i$

基本假设　　$f_0, f_1, \cdots, f_m$ 均为具有连续二阶导数的凸函数

$A \in R^{p \times n}$ 为行满秩矩阵

存在最优解 x^*

存在 $x \in D = \bigcap_{0 \le i \le m} \mathrm{dom}\, f_i$ 满足 $f_i(x) < 0,\ i = 1, \ldots, m,\ Ax = b$

17.1. Ineq. Const. Minimization Problems

在以上假设下，根据 Slater 定理，$\left(x^*,\lambda^*,\nu^*\right)$是原对偶问题最优解的充要条件是它们同时满足以下 KKT 方程

$$
\begin{aligned}
& f_i\left(x^*\right) \le 0, \quad \forall 1 \le i \le m \\
& Ax^* = b \\
& \lambda_i^* f_i\left(x^*\right) = 0, \quad \forall 1 \le i \le m \\
& \lambda^* \ge 0 \\
& \nabla f_0\left(x^*\right) + \sum_{i=1}^{m} \lambda_i^* \nabla f_i\left(x^*\right) + A^T \nu^* = 0
\end{aligned}
$$

17.2. Barrier Method

非正实数集合的指示函数及其（可导）近似函数

$$I_{-}(u)=\begin{cases}0 & \forall u\le 0\\ \infty & \forall u>0\end{cases}$$

近似函数 the log barrier

$$\hat{I}_{-}(u)=-\frac{1}{t}\log(-u)$$

where $t>0$ is a large number

This approximation is more accurate for larger t. But for any value of t, the log barrier approaches ∞ if any $t>0$

17.2. Barrier Method

指示函数及其近似函数的图像

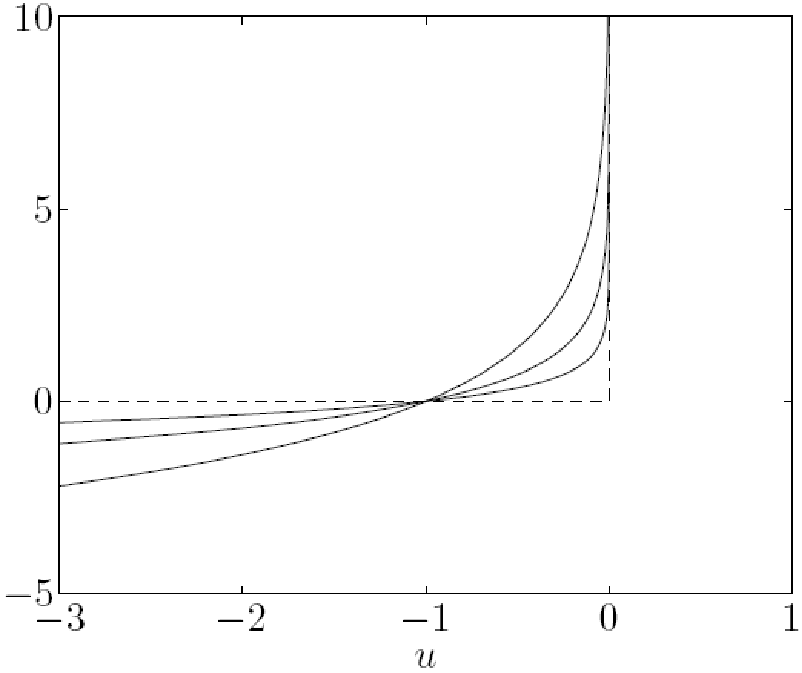


Figure 1 点划线表示函数$I_-(u)$，实曲线分别表示$t = 0.5$，1，2所对应的$\widehat{I}_-(u) = -(1/t)\log(-u)$. 对应于$t = 2$的曲线给出最好的近似.

近似函数是负实轴上任意阶可导的凸函数

对数障碍函数及其偏导数

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

$$\nabla \phi(x) = \sum_{i=1}^{m} \frac{1}{-f_i(x)} \nabla f_i(x)$$

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

近似优化问题

$$\min\left\{ f_0(x) \middle| \text{ s.t. } f_i(x) \le 0, i = 1, \cdots, m, Ax = b \right\}$$

$$\Leftrightarrow \quad \min\left\{ f_0(x) + \sum_{i=1}^{m} I_-\left(f_i(x)\right) \middle| \text{ s.t. } Ax = b \right\}$$

$$\approx \quad \min\left\{ f_0(x) + \frac{1}{t}\phi(x) \middle| \text{ s.t. } Ax = b \right\}$$

$$\Leftrightarrow \quad \min\left\{ t f_0(x) + \phi(x) \middle| \text{ s.t. } Ax = b \right\}$$

17.2. Barrier Method

对任意 $t > 0$，用 $x^*(t)$ 表示上述近似优化问题的最优解，即

$$tf_0\left(x^*(t)\right) + \phi\left(x^*(t)\right) = \min\left\{tf_0(x) + \phi\left(x\right) \middle| \text{s.t. } Ax = b\right\}$$

我们将 $\left\{x^*(t),\ t > 0\right\}$ 定义为该优化问题中心路径 Central Path

显然，我们期望 $t \to +\infty$ 时，$x^*(t) \to x^*$

Why don't we just set t to be some huge value, and solve the above problem? Directly seek solution at end of central path?

Problem is that this is seriously ineffcient in practice. Much more effcient to traverse the central path, as we will see

17.2. Barrier Method

An important special case: barrier problem for a linear program:

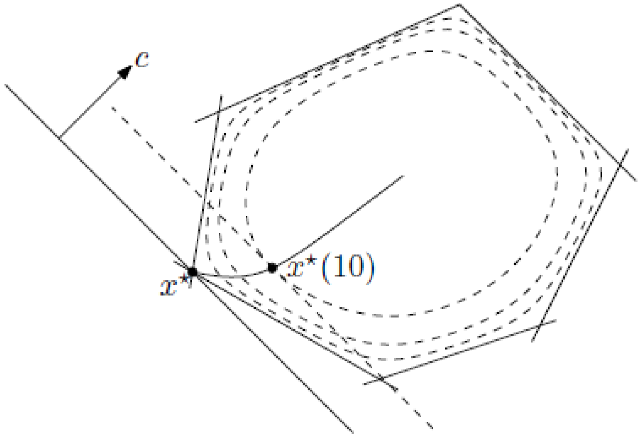
$$\min_x \; tc^T x - \sum_{i=1}^{m} \log(e_i - d_i^T x)$$

The barrier function corresponds to polyhedral constraint $Dx \leq e$

Gradient optimality condition:

$$0 = tc - \sum_{i=1}^{m} \frac{1}{e_i - d_i^T x^\star(t)} d_i$$

This means that gradient $\nabla\phi(x^\star(t))$ must be parallel to $-c$, i.e., hyperplane $\{x : c^T x = c^T x^\star(t)\}$ lies tangent to contour of ϕ at $x^\star(t)$



(From B & V page 565)

17.2. Barrier Method

中心路径 $x^*(t)$ 一定满足某个如下的 KKT 条件

$$t\nabla f_0\left(x^*(t)\right) + \nabla\phi\left(x^*(t)\right) + A^T w = 0$$

$$Ax^*(t) = b$$

$$\Leftrightarrow \quad \begin{aligned} & t\nabla f_0\left(x^*(t)\right) + \sum_{i=1}^{m} \frac{1}{-f_i\left(x^*(t)\right)} \nabla f_i\left(x^*(t)\right) + A^T w = 0 \\ & Ax^*(t) = b \end{aligned}$$

$w \in R^m$，我们实际上不用去关心原始问题的对偶问题的最优解，因为对于给定的 $x^*(t)$ 和 w，我们可以直接令

$$\lambda_i^*(t) = -\frac{1}{tf_i\left(x^*(t)\right)}, i = 1, \cdots, m, \quad \nu^*(t) = \frac{w}{t}$$

则 $\lambda_i^*(t), \nu^*(t)$ 则为原始问题的对偶问题的可行解

17.2. Barrier Method

We claim $u^\star(t), v^\star(t)$ are dual feasible for original problem, whose Lagrangian is

$$L(x, u, v) = f(x) + \sum_{i=1}^{m} u_i h_i(x) + v^T(Ax - b)$$

Why?

- Note that $u_i^\star(t) > 0$ since $h_i(x^\star(t)) < 0$ for all $i = 1, \ldots, m$
- Further, the point $(u^\star(t), v^\star(t))$ lies in domain of Lagrange dual function $g(u, v)$, since by definition

$$\nabla f(x^\star(t)) + \sum_{i=1}^{m} u_i(x^\star(t)) \nabla h_i(x^\star(t)) + A^T v^\star(t) = 0$$

  That is, $x^\star(t)$ minimizes Lagrangian $L(x, u^\star(t), v^\star(t))$ over x, so $g(u^\star(t), v^\star(t)) > -\infty$

This allows us to bound suboptimality of $f_i\left(x^*(t)\right)$, with respect to original problem, via the duality gap.

令 $$\lambda_i^*(t) = -\frac{1}{tf_i\left(x^*(t)\right)},\ i = 1,\cdots,m, \quad \nu^*(t) = \frac{w}{t}$$

$\Rightarrow$
$$\nabla f_0(x^*(t)) + \sum_{i=1}^{m} \lambda_i^*(t) \nabla f_i(x^*(t)) + A^T \nu^*(t) = 0, \quad Ax^*(t) - b = 0$$
$$-\lambda_i^* f_i(x^*(t)) - \frac{1}{t} = 0,\ 1 \le i \le m$$

$\Rightarrow$
$$g\left(\lambda^*(t), \nu^*(t)\right) = \min_{x \in D} f_0(x) + \sum_{i=1}^{m} \lambda_i^*(t) f_i(x) + (Ax - b)^T \nu^*(t)$$
$$= f_0\left(x^*(t)\right) + \sum_{i=1}^{m} \lambda_i^*(t) f_i\left(x^*(t)\right) + \left(Ax^*(t) - b\right)^T \nu^*(t) = f_0\left(x^*(t)\right) - \frac{m}{t}$$

$\Rightarrow$ $$f_0\left(x^*(t)\right) - p^* \le \frac{m}{t}$$ $\Rightarrow$ $$\lim_{t \to \infty} f_0\left(x^*(t)\right) = p^*$$

This will be very useful as a stopping criterion

17.2. Barrier Method

We can now reinterpret central path $(x^\star(t), u^\star(t), v^\star(t))$ as solving the perturbed KKT conditions:

$$\nabla f(x) + \sum_{i=1}^{m} u_i \nabla h_i(x) + A^T v = 0$$

$$u_i \cdot h_i(x) = -1/t, \quad i = 1, \ldots, m$$

$$h_i(x) \le 0, \quad i = 1, \ldots, m, \quad Ax = b$$

$$u_i \ge 0, \quad i = 1, \ldots, m$$

Only difference between these and actual KKT conditions for our original problem is second line: these are replaced by

$$u_i \cdot h_i(x) = 0, \quad i = 1, \ldots, m$$

i.e., complementary slackness, in actual KKT conditions

17.2. Barrier Method

障碍方法总结

算法条件：已知 x^0 满足 $f_i\left(x^0\right)<0, 1\le i\le m$

1）给定 $t^0>0$， $\mu>1$， $\varepsilon>0$， $k=0$

2）以 x^k 为初始点用等式约束 Newton 法解以下问题得到 x^{k+1}

$$\min\left\{t^k f_0(x)+\phi\left(x\right)\middle|\text{s.t. } Ax=b\right\}$$

This step is called a centering step (since it brings x^{k+1} onto the central path)

3）如果 $\frac{m}{t^k}\le\varepsilon$，停止，否则令 $t^{k+1}=\mu t^k$，用 $k+1$ 替换 k，回 2）

17.2. Barrier Method

Considerations:

- Choice of μ: if μ is too small, then many outer iterations might be needed; if μ is too big, then Newton's method (each centering step) might take many iterations
- Choice of $t^{(0)}$: if $t^{(0)}$ is too small, then many outer iterations might be needed; if $t^{(0)}$ is too big, then the first Newton solve (first centering step) might require many iterations

Fortunately, the performance of the barrier method is often quite robust to the choice of μ and $t^{(0)}$ in practice

(However, note that the appropriate range for these parameters is scale dependent)

17.2. Barrier Method

Another 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 $F_i(x) = (1/f_i(x))\nabla f_i(x)$

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

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

17.2. Barrier Method

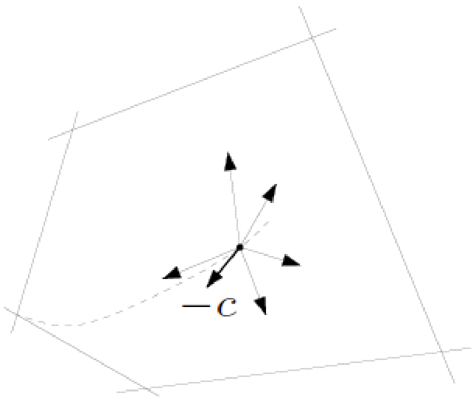
example

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

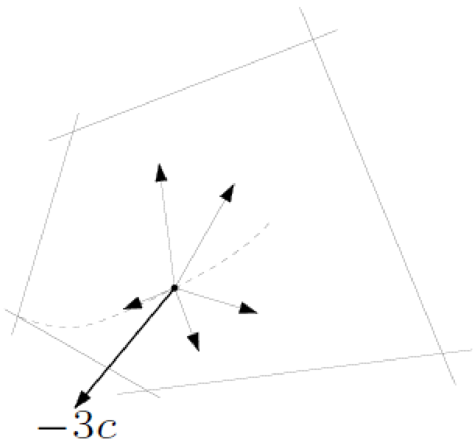
- objective force field is constant: $F_0(x) = -tc$
- constraint force field decays as inverse distance to constraint hyperplane:

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

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



$t = 1$



$t = 3$

17.2. Barrier Method

例、不等式约束的线性规划问题 $\min\left\{c^T x \middle| \text{s.t. } \bar{A}x \le \bar{b}\right\}$， $\bar{A} \in R^{100\times 50}$

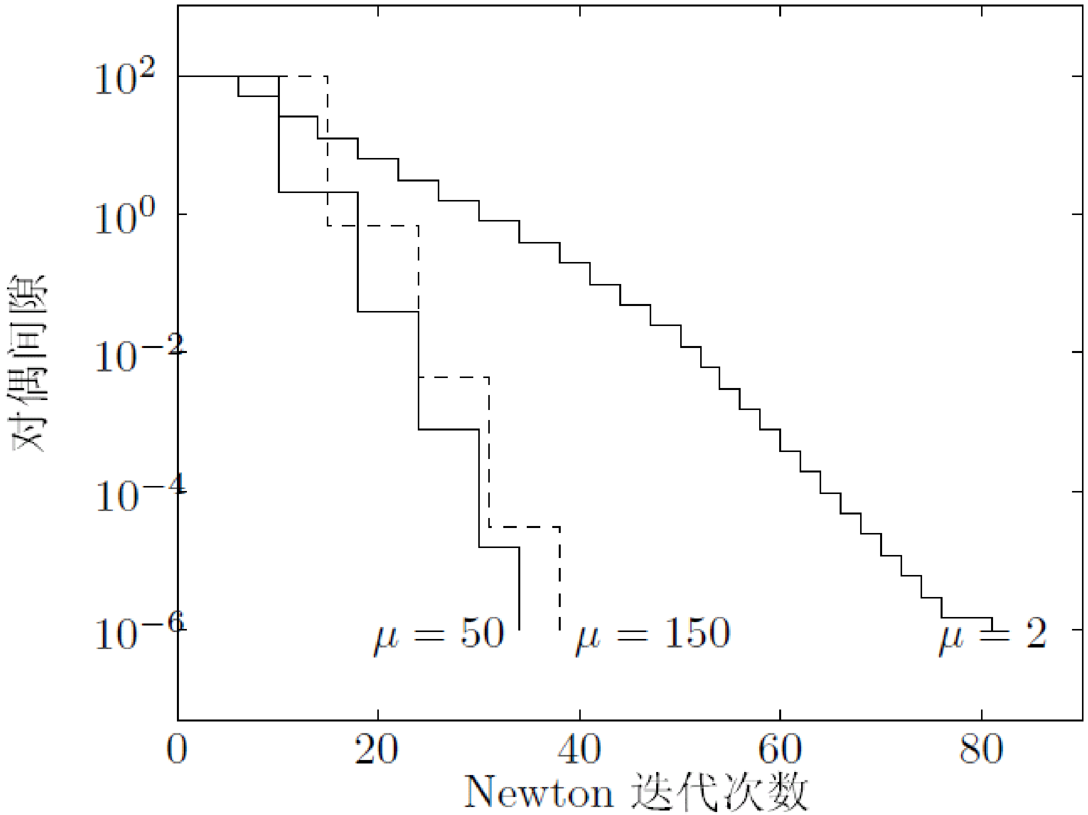


图 11.4 用障碍方法求解一个小规模线性规划问题，迭代过程中对偶间隙和累计 Newton 迭代次数之间的关系。所给出的三条曲线对应于参数 μ 的三个数值：2，50 和 150。每种情况下的对偶间隙均显示近似线性的收敛性。

17.2. Barrier Method

Newton 迭代次数和参数 μ 之间的关系

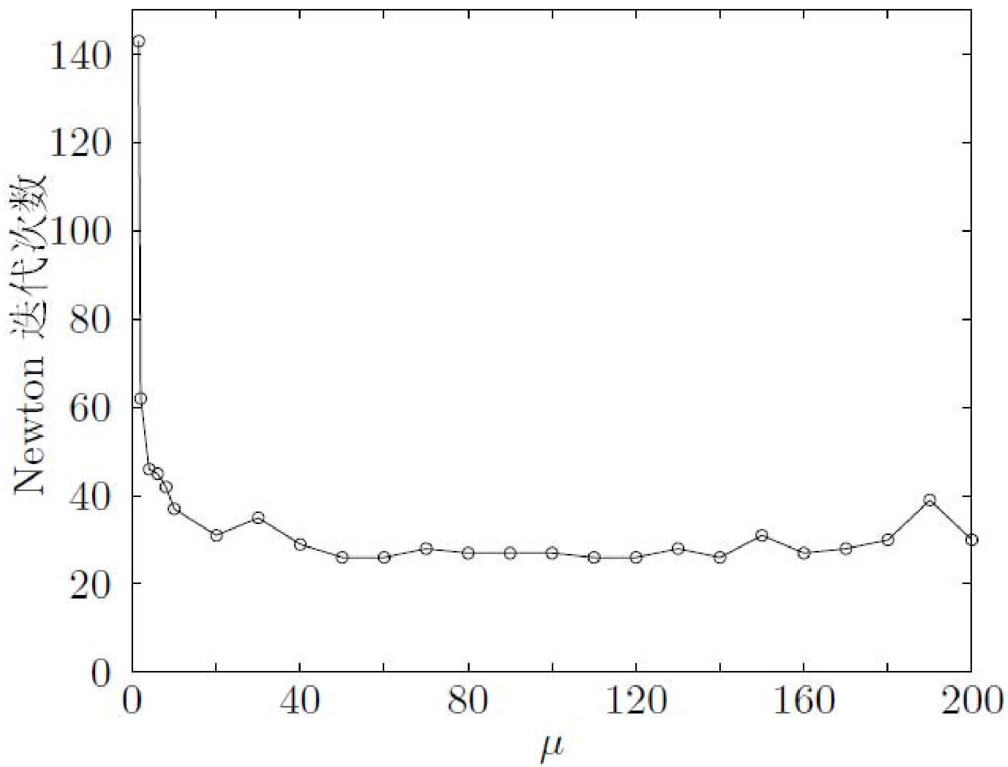


图 11.5 对一个小规模线性规划问题选择不同 μ 值的效果。纵轴表示将对偶间隙从 100 减少到 10^{-3} 所需要的 Newton 迭代次数，横轴表示 μ 值。图像表明当 μ 大于 3 或附近的数值以后障碍方法就可以取得较好的效果，但在很大范围内对 μ 值并不敏感。

17.2. Barrier Method

例、凸几何规划

$$\min\left\{\log\left(\sum_{k=1}^{5}\exp\left(a_{0k}^{T}x+b_{0k}\right)\right)\middle|\text{s.t. }\log\left(\sum_{k=1}^{5}\exp\left(a_{ik}^{T}x+b_{ik}\right)\right)\le 0, 1\le i\le m\right\},\quad x\in R^{50},\quad m=100$$

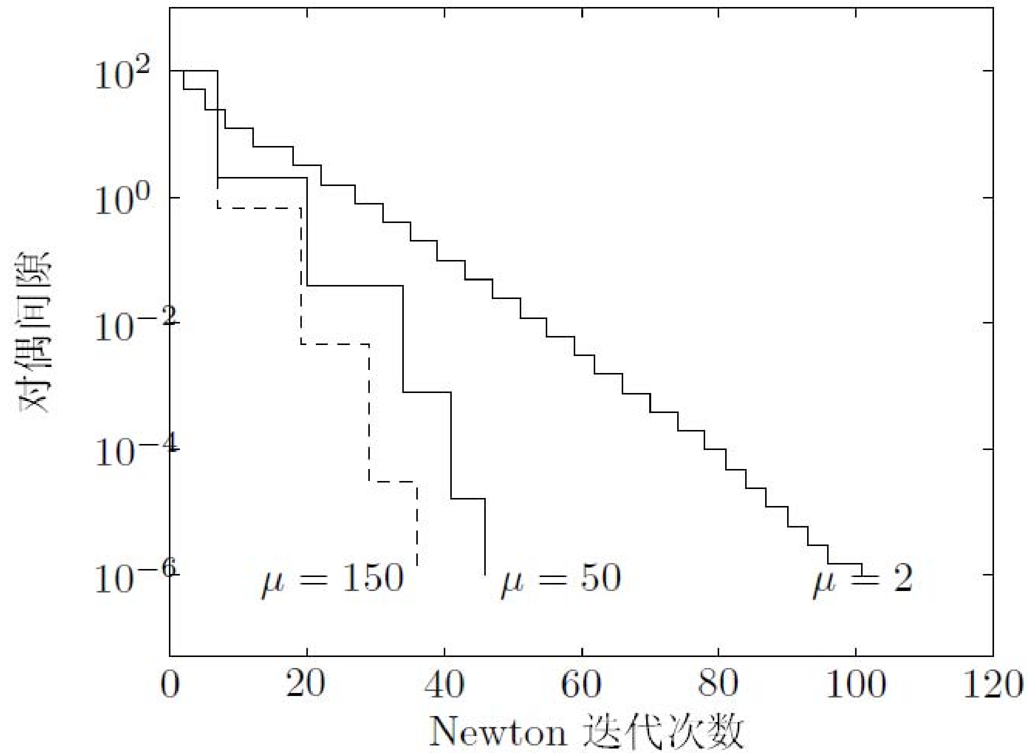


图 11.6 用障碍方法求解一个小规模几何规划规划问题。曲线显示对偶间隙和累计 Newton 迭代次数之间的关系。对偶间隙依然近似线性收敛。

17.2. Barrier Method

例、标准线性规划问题 $\min\left\{c^T x \middle| \text{s.t. } Ax = b, x \geq 0\right\}$， $A \in R^{m\times n}$

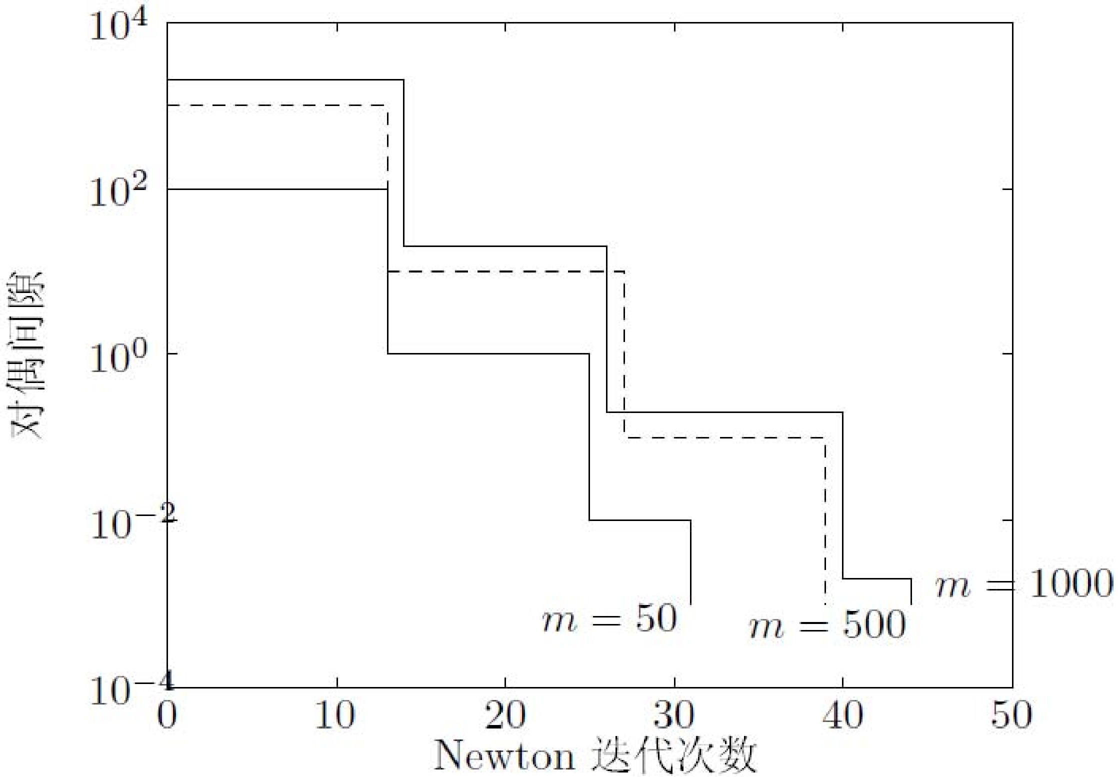


图 11.7　用障碍方法求解三个随机产生的不同维数的线性规划问题。曲线表示对偶间隙和累计 Newton 迭代次数之间的关系。每个问题的变量个数为 $n = 2m$。对偶间隙依然近似线性收敛。问题规模增加时所需要的 Newton 迭代次数仅有少量增加。

17.2. Barrier Method

对每个 n 随机产生 100 个问题的计算结果

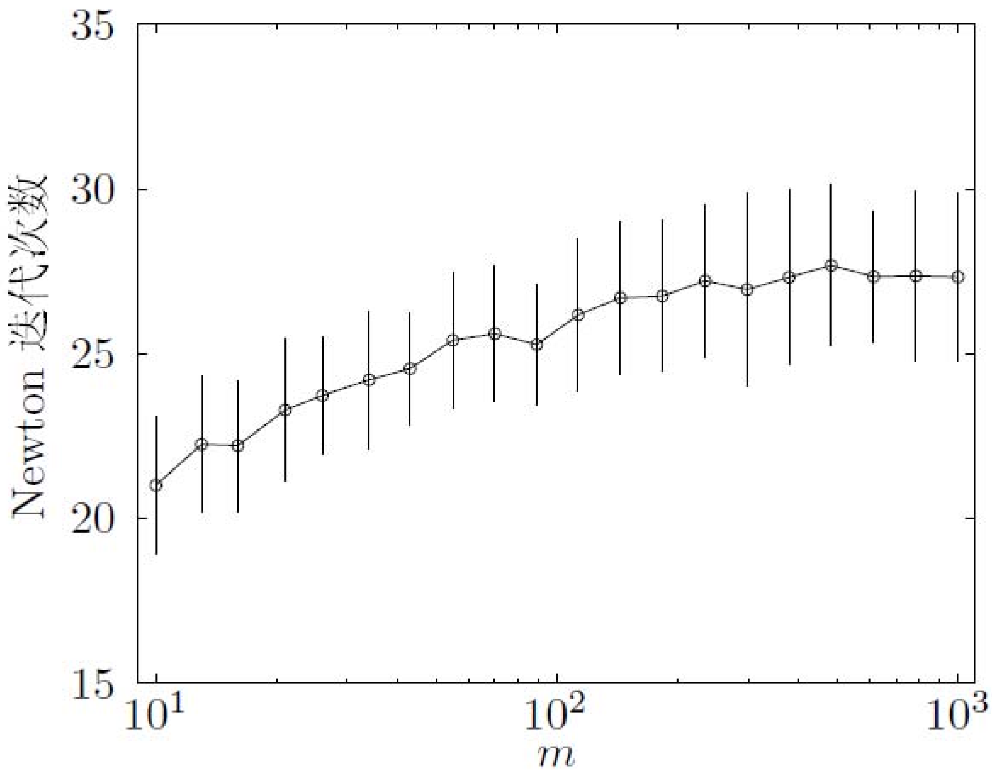


图 11.8 对于不同维数，求解 $n = 2m$ 的 100 个随机产生的线性规划问题所需要的 Newton 迭代次数的均值。对每个 m，过均值的误差线段表示标准差。当问题维数的变化比值为 100 : 1 时，所需要的 Newton 迭代次数的增量非常小。

17.2. Barrier Method

Assume that we solve the centering steps exactly. The following result is immediate

Theorem: The barrier method after k centering steps satisfies

$$f(x^{(k)}) - f^\star \leq \frac{m}{\mu^k t^{(0)}}$$

In other words, to reach a desired accuracy level of ϵ, we require

$$\frac{\log(m/(t^{(0)}\epsilon))}{\log \mu}$$

centering steps with the barrier method (plus initial centering step)

17.2. Barrier Method

How many Newton iterations do we need?

Informally, due to careful central path traversal, in each centering step, Newton is already in quadratic convergence phase, so takes nearly constant number of iterations

This can be formalized under self-concordance. Suppose:

- The function $tf + \phi$ is self-concordant
- Our original problem has bounded sublevel sets

Then we can terminate each Newton solve at appropriate accuracy, and the total number of Newton iterations is still $O(\log(m/(t^{(0)}\epsilon))$ (where constants do not depend on problem-specific conditioning). See Chapter 11.5 of B & V

Importantly, $tf + \phi = tf - \sum_{i=1}^m \log(-h_i)$ is self-concordant when f, h_i are all linear or quadratic. This covers all LPs, QPs, QCQPs

17.2. Barrier Method

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 $\lambda = \lambda^\star(t)$, $\nu = \nu^\star(t)$):

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

17.2. Barrier Method

total number of Newton iterations (excluding first centering step)

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

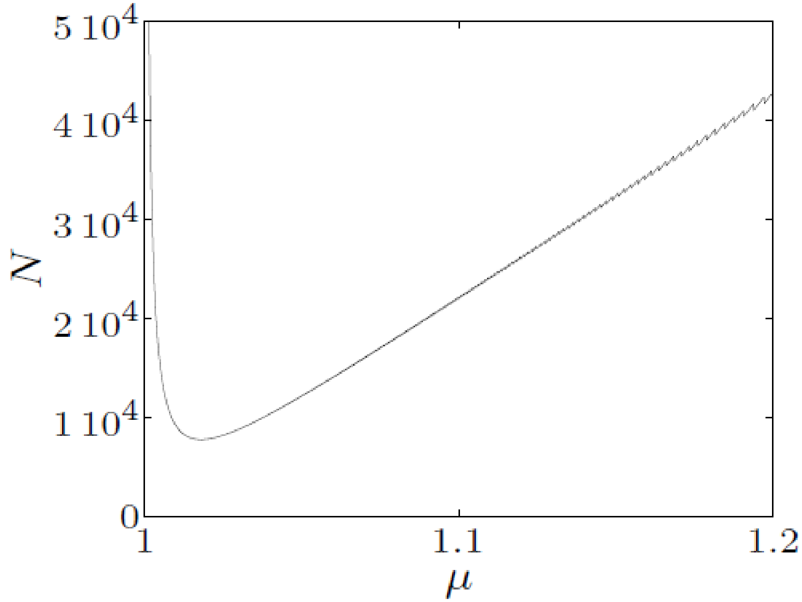


figure shows N for typical values of γ, c,

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

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

17.2. Barrier Method

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 ($\mu = 10, \ldots, 20$)

17.2. Barrier Method

We have implicitly assumed that we have a strictly feasible point for the first centering step, i.e., for computing $x^{(0)} = x^\star$, solution of barrier problem at $t = t^{(0)}$. This is x such that

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

How to find such a feasible x? By solving

$$\begin{aligned} \min_{x,s} \quad & s \\ \text{subject to} \quad & h_i(x) \leq s, \; i = 1, \ldots, m \\ & Ax = b \end{aligned}$$

The goal is for s to be negative at the solution. This is known as a feasibility method. We can apply the barrier method to the above problem, since it is easy to find a strictly feasible starting point

17.2. Barrier Method

Note that we do not need to solve this problem to high accuracy. Once we find a feasible (x, s) with $s < 0$, we can terminate early

An alternative is to solve the problem

$$\begin{array}{ll} \min\limits_{x,s} & 1^T s \\ \text{subject to} & h_i(x) \le s_i, \; i = 1, \ldots, m \\ & Ax = b, \; s \ge 0 \end{array}$$

Previously s was the maximum infeasibility across all inequalities. Now each inequality has own infeasibility variable s_i, $i = 1, \ldots, m$

One advantage: when the original system is infeasible, the solution of the above problem will be informative. The nonzero entries of s will tell us which of the constraints cannot be satisfied

17.2. Barrier Method

总结可行性和阶段 1 方法

可行性问题： $f_i(x) \le 0, i=1,\cdots,m, Ax=b$ 是否有解

极小化最大不可行值

$$\min\left\{s \,\middle|\, \text{s.t. } f_i(x) \le s, i=1,\cdots,m, Ax=b\right\}$$

极小化不可行值之和

$$\min\left\{1^T s \,\middle|\, \text{s.t. } f_i(x) \le s_i, i=1,\cdots,m, Ax=b, s \ge 0\right\}$$

根据最优目标值判断是否可行，并在可行情况下得初始解

17.2. Barrier Method

两种阶段 1 方法的比较

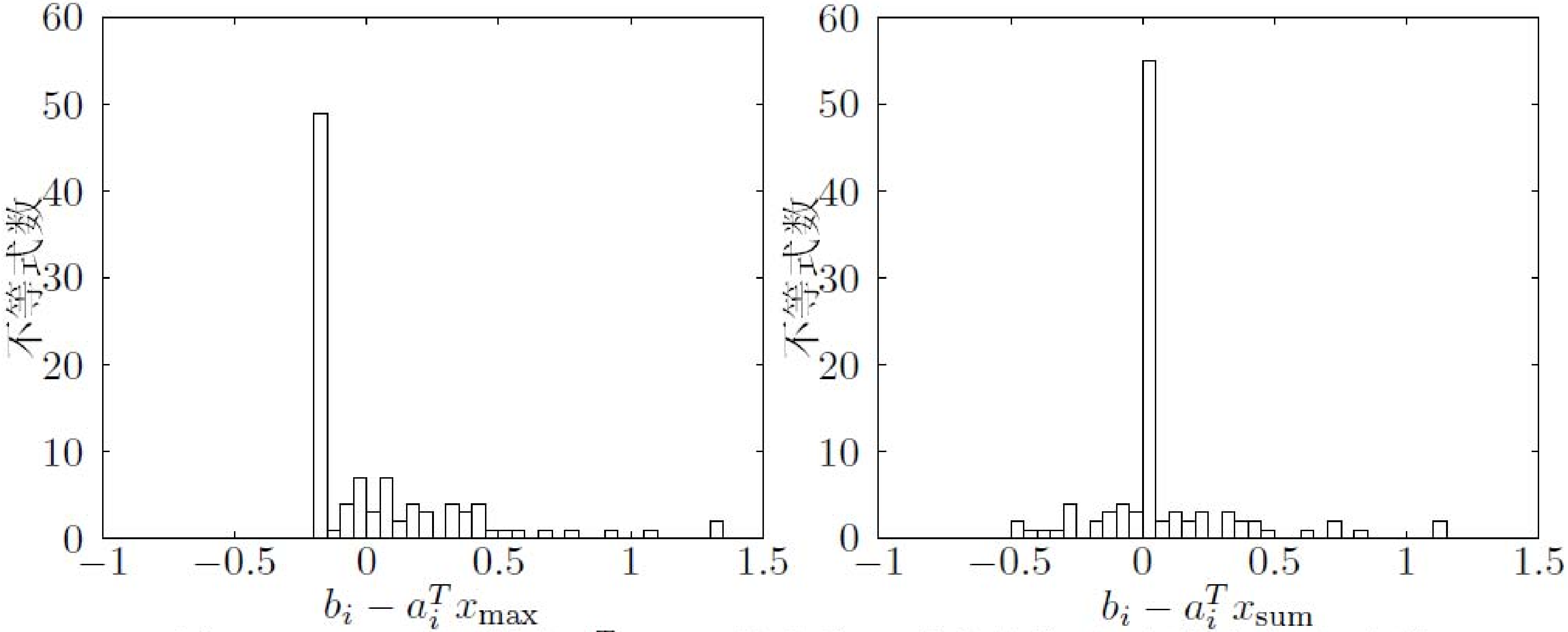
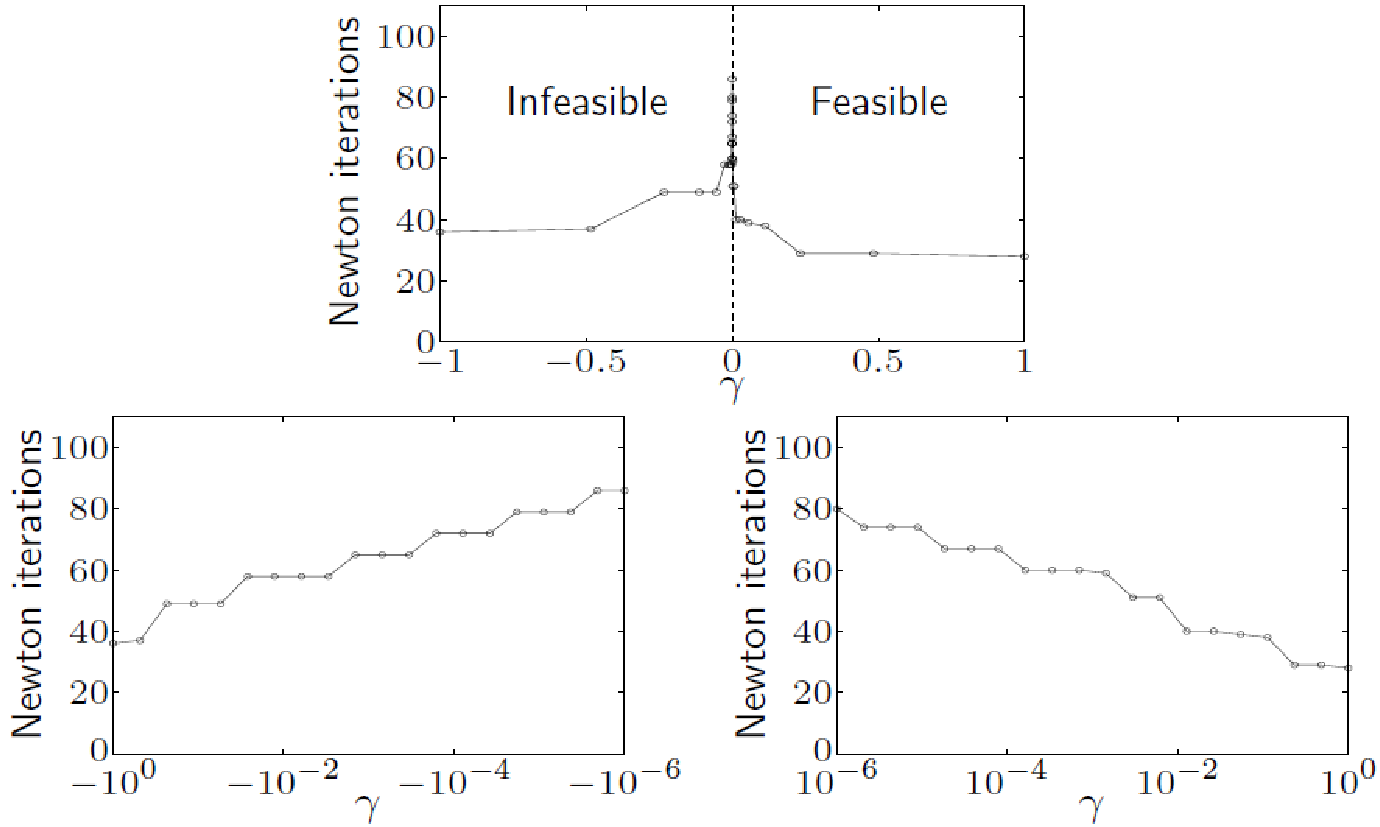


图 11.9 100 个不等式 $a_i^T x \leqslant b_i$ 构成的 50 个变量的不可行集合中不可行值 $b_i - a_i^T x$ 的分布情况。左边图像使用的向量 $x_{\max}$ 是基本的阶段 1 方法给出的解，它满足 100 个不等式中的 39 个。右边图像使用的向量 x_{sum} 是极小化不可行值之和给出的解，它满足 100 个不等式中的 79 个。

17.2. Barrier Method

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 $\log(1/|\gamma|)$

17.2. Barrier Method

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 $K_i \in \mathbf{R}^{k_i}$
- f_i twice continuously differentiable
- $A \in \mathbf{R}^{p \times n}$ 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

17.2. Barrier Method

Generalized logarithm for proper cone

$\psi : \mathbf{R}^q \to \mathbf{R}$ is generalized logarithm for proper cone $K \subseteq \mathbf{R}^q$ 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 $K = \mathbf{R}^n_+$: $\psi(y) = \sum_{i=1}^n \log y_i$, with degree $\theta = n$
- positive semidefinite cone $K = \mathbf{S}^n_+$:

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

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

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

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

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

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

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

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

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

17.2. Barrier Method

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 $w \in \mathbf{R}^p$,

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

($Df_i(x) \in \mathbf{R}^{k_i \times n}$ 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 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

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

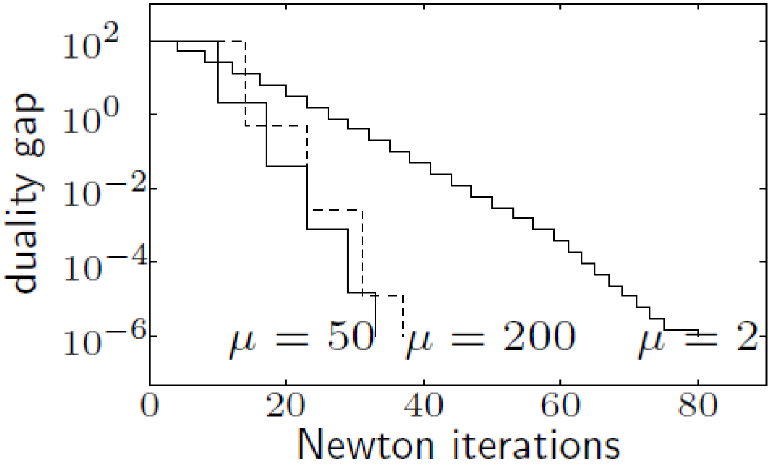
- dual point on central path: $Z^\star(t) = -(1/t)F(x^\star(t))^{-1}$ is feasible for

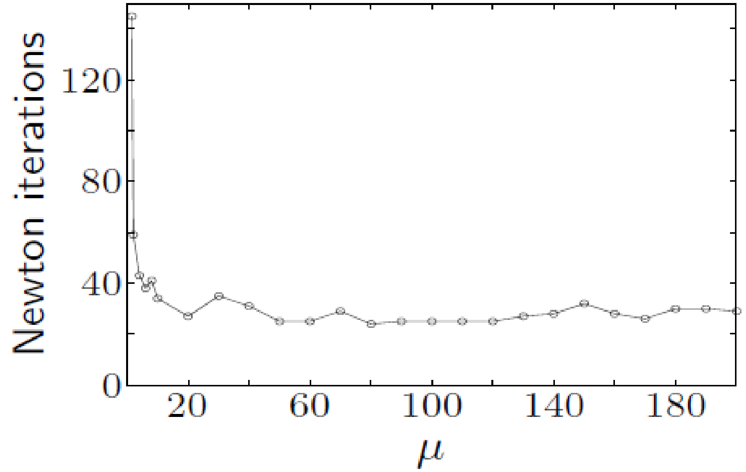
$$\begin{array}{ll} \text{maximize} & \mathbf{tr}(GZ) \\ \text{subject to} & \mathbf{tr}(F_i Z) + c_i = 0, \quad i = 1, \ldots, n \\ & Z \succeq 0 \end{array}$$

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

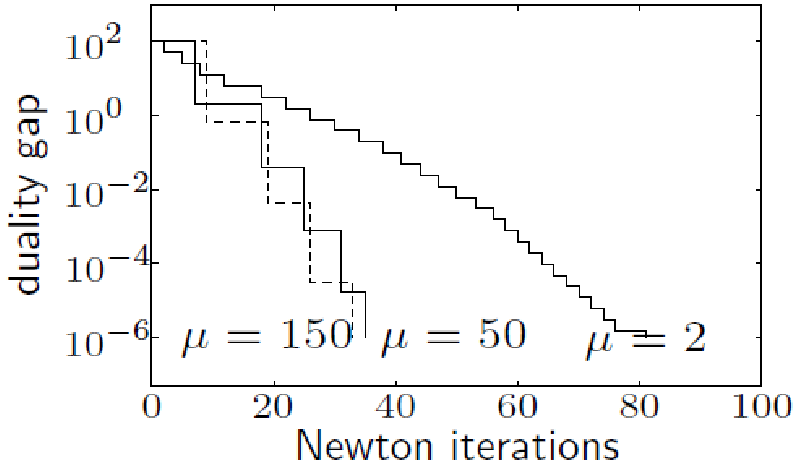
17.2. Barrier Method

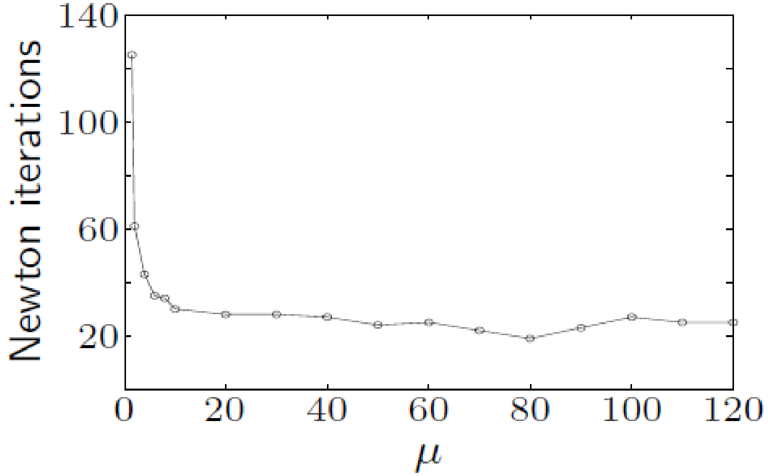
second-order cone program (50 variables, 50 SOC constraints in $\mathbf{R}^6$)





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



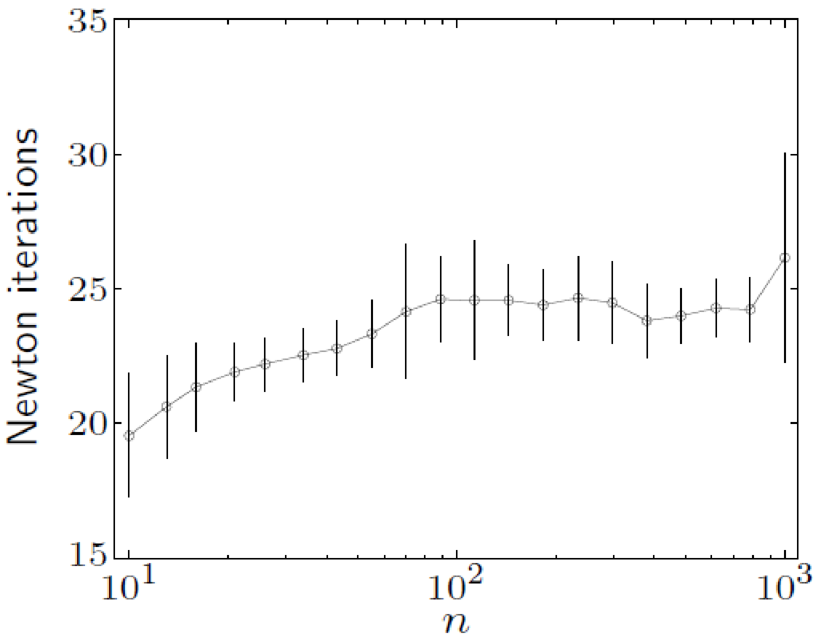


17.2. Barrier Method

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



17.3. 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
- Primal-dual interior-point methods are less intuitive ...

17.3. Primal-Dual Interior-Point Methods

基本思想

障碍方法求解不等式约束问题的基本步骤：

1）给定$t>0$求解下述问题得到中心点$x^*(t)$

$$
\begin{aligned}
&\min\ tf_0(x)-\sum_{i=1}^{m}\log\left(-f_i(x)\right)\\
&\text{s.t.}\quad Ax=b
\end{aligned}
$$

2）增加t直至$\frac{m}{t}\le\varepsilon$，此时$f_0\left(x^*(t)\right)-p^*\le\frac{m}{t}\le\varepsilon$

17.3. Primal-Dual Interior-Point Methods

第一个步骤等价于解下述 KKT 条件

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

令 $\lambda_i = -\dfrac{1}{tf_i(x)}, \forall i, \quad \nu = \dfrac{w}{t}$，又等价于解下述修改的 KKT 条件

$$
\begin{aligned}
& \nabla f_0(x) + \sum_{i=1}^{m} \lambda_i \nabla f_i(x) + A^T \nu = 0 \\
& -\lambda_i f_i(x) - \frac{1}{t} = 0, \ 1 \le i \le m \\
& Ax - b = 0
\end{aligned}
$$

17.3. Primal-Dual Interior-Point Methods

Can view this as a nonlinear system of equations, written as

$$r(x,u,v) = \begin{pmatrix} \nabla f(x) + Dh(x)^T u + A^T v \\ -\mathrm{diag}(u)h(x) - (1/t)1 \\ Ax - b \end{pmatrix} = 0$$

where

$$h(x) = \begin{pmatrix} h_1(x) \\ \dots \\ h_m(x) \end{pmatrix}, \quad Dh(x) = \begin{bmatrix} \nabla h_1(x)^T \\ \dots \\ \nabla h_m(x)^T \end{bmatrix}$$

Newton's method, recall, is generally a root-finder for a nonlinear system $F(y) = 0$. Approximating $F(y + \Delta y) \approx F(y) + DF(y)\Delta y$ leads to

$$\Delta y = -(DF(y))^{-1} F(y)$$

What happens if we apply this to $r(x,u,v) = 0$ above?

17.3. Primal-Dual Interior-Point Methods

Approach 1: from middle equation (relaxed comp slackness), note that $u_i = -1/(th_i(x))$, $i = 1, \ldots, m$. So after eliminating u, we get

$$r(x, v) = \begin{pmatrix} \nabla f(x) + \sum_{i=1}^{m} (-\frac{1}{th_i(x)}) \nabla h_i(x) + A^T v \\ Ax - b \end{pmatrix} = 0$$

Thus the Newton root-finding update $(\Delta x, \Delta v)$ is determined by

$$\begin{bmatrix} H_{\text{bar}}(x) & A^T \\ A & 0 \end{bmatrix} \begin{pmatrix} \Delta x \\ \Delta v \end{pmatrix} = -r(x, v)$$

where $H_{\text{bar}}(x) =$
$\nabla^2 f(x) + \sum_{i=1}^{m} \frac{1}{th_i(x)^2} \nabla h_i(x) \nabla h_i(x)^T + \sum_{i=1}^{m} (-\frac{1}{th_i(x)}) \nabla^2 h_i(x)$

This is just the KKT system solved by one iteration of Newton's method for minimizing the barrier problem

17.3. Primal-Dual Interior-Point Methods

Approach 2: directly apply Newton root-finding update, without eliminating u. Introduce notation

$$
\begin{aligned}
r_{\text{dual}} &= \nabla f(x) + Dh(x)^T u + A^T v \\
r_{\text{cent}} &= -\text{diag}(u)h(x) - (1/t)t \\
r_{\text{prim}} &= Ax - b
\end{aligned}
$$

called the dual, central, and primal residuals at $y = (x, u, v)$. Now root-finding update $\Delta y = (\Delta x, \Delta u, \Delta v)$ is given by

$$
\begin{bmatrix} H_{\text{pd}}(x) & Dh(x)^T & A^T \\ -\text{diag}(u)Dh(x) & -\text{diag}(h(x)) & 0 \\ A & 0 & 0 \end{bmatrix} \begin{pmatrix} \Delta x \\ \Delta u \\ \Delta v \end{pmatrix} = -\begin{pmatrix} r_{\text{dual}} \\ r_{\text{cent}} \\ r_{\text{prim}} \end{pmatrix}
$$

where $H_{\text{pd}}(x) = \nabla^2 f(x) + \sum_{i=1}^m u_i \nabla^2 h_i(x)$

17.3. Primal-Dual Interior-Point Methods

将上述 Approach 1 和 Approach 2 简记为 v1 和 v2，我们可以知道

- In v2, update directions for the primal and dual variables are inexorably linked together
- Also, v2 and v1 leads to different (nonequivalent) updates
- As we saw, one iteration of v1 is equivalent to inner iteration in the barrier method
- And v2 defines a new method called primal-dual interior-point method, that we will flesh out shortly
- One complication: in v2, the dual iterates are not necessarily feasible for the original dual problem ...

17.3. Primal-Dual Interior-Point Methods

For barrier method, we have simple duality gap: m/t, since we set $u_i = -1/(th_i(x))$, $i = 1, \ldots, m$ and saw this was dual feasible

For primal-dual interior-point method, we can construct surrogate duality gap:

$$\eta = -h(x)^T u = -\sum_{i=1}^{m} u_i h_i(x)$$

This would be a bonafide duality gap if we had feasible points, i.e., $r_{\text{prim}} = 0$ and $r_{\text{dual}} = 0$, but we don't, so it's not

What value of parameter t does this correspond to in perturbed KKT conditions? This is $t = m/\eta$

17.3. Primal-Dual Interior-Point Methods

令

$$\eta(x,\lambda) = -\sum_{i=1}^{m} \lambda_i f_i(x)$$

则

$$t = \frac{m}{\eta(x,\lambda)}, \quad \eta(x,\lambda) = \frac{m}{t}$$

障碍方法的收敛条件 $\frac{m}{t} \le \varepsilon$ 等价于

$$\eta(x,\lambda) \le \varepsilon$$

由于 $t = \frac{m}{\eta(x,\lambda)}$，每次更新 x,λ 后即可更新 t

17.3. Primal-Dual Interior-Point Methods

Putting it all together, we now have our primal-dual interior-point method. Start with $x^{(0)}$ such that $h_i(x^{(0)}) < 0$, $i = 1, \ldots, m$, and $u^{(0)} > 0$, $v^{(0)}$. Define $\eta^{(0)} = -h(x^{(0)})^T u^{(0)}$. We fix $\mu > 1$, repeat for $k = 1, 2, 3 \ldots$

- Define $t = \mu m / \eta^{(k-1)}$
- Compute primal-dual update direction Δy
- Use backtracking to determine step size s
- Update $y^{(k)} = y^{(k-1)} + s \cdot \Delta y$
- Compute $\eta^{(k)} = -h(x^{(k)})^T u^{(k)}$
- Stop if $\eta^{(k)} \leq \epsilon$ and $(\|r_{\text{prim}}\|_2^2 + \|r_{\text{dual}}\|_2^2)^{1/2} \leq \epsilon$

Note that we stop based on surrogate duality gap, and approximate feasibility. (Line search maintains $h_i(x) < 0$, $u_i > 0$, $i = 1, \ldots, m$)

17.3. Primal-Dual Interior-Point Methods

At each step, must ensure we arrive at $y^+ = y + s\Delta y$, i.e.,

$$x^+ = x + s\Delta x, \quad u^+ = u + s\Delta u, \quad v^+ = v + s\Delta v$$

that maintains both $h_i(x) < 0$, and $u_i > 0$, $i = 1, \ldots, m$

A multi-stage backtracking line search for this purpose: start with largest step size $s_{\max} \le 1$ that makes $u + s\Delta u \ge 0$:

$$s_{\max} = \min\left\{1,\ \min\{-u_i/\Delta u_i : \Delta u_i < 0\}\right\}$$

Then, with parameters $\alpha, \beta \in (0, 1)$, we set $s = 0.999 s_{\max}$, and

- Let $s = \beta s$, until $h_i(x^+) < 0$, $i = 1, \ldots, m$
- Let $s = \beta s$, until $\|r(x^+, u^+, v^+)\|_2 \le (1 - \alpha s)\|r(x, u, v)\|_2$

17.3. Primal-Dual Interior-Point Methods

保证收敛的二次规划原对偶内点法（Jacek Gondio, 2012）

原问题 $\min\left\{c^T x+\frac{1}{2}x^T Qx \,\middle|\, \text{s.t. } Ax=b, x\geq 0\right\}$，其中 $A\in R^{m\times n}$ 为行满秩矩阵， $Q\in S_+^n$ （ $Q=0$ 就成为标准线性规划问题）

由于 $f_0(x)=c^T x+\frac{1}{2}x^T Qx$， $f_i(x)=-x_i$， $i=1,2,\cdots,n$，则 $\nabla^2 f_0(x)=Q$， $\nabla^2 f_i(x)=0$， $i=1,2,\cdots,n$， $Df(x)^T=-I$， Newton 方向为：

$$\begin{pmatrix}\Delta x\\ \Delta\lambda\\ \Delta\nu\end{pmatrix}=-\begin{pmatrix}Q & -I & A^T\\ \mathrm{diag}(\lambda) & \mathrm{diag}(x) & 0\\ A & 0 & 0\end{pmatrix}^{-1}\begin{pmatrix}c+Qx-\lambda+A^T\nu\\ \mathrm{diag}(\lambda)x-(1/t)\vec{1}\\ Ax-b\end{pmatrix}$$

直线搜索： $t=0.99\max\left\{\tau \,\middle|\, \text{s.t. } x+\tau\Delta x\geq 0, \lambda+\tau\Delta\lambda\geq 0\right\}$ ！！！

17.3. Primal-Dual Interior-Point Methods

Recall the standard form LP:

$$
\begin{aligned}
\min_{x} \quad & c^T x \\
\text{subject to} \quad & Ax = b \\
& x \geq 0
\end{aligned}
$$

for $c \in \mathbb{R}^n$, $A \in \mathbb{R}^{m \times n}$, $b \in \mathbb{R}^m$. Its dual is:

$$
\begin{aligned}
\max_{u,v} \quad & b^T v \\
\text{subject to} \quad & A^T v + u = c \\
& u \geq 0
\end{aligned}
$$

(This is not a bad thing to memorize)

17.3. Primal-Dual Interior-Point Methods

The points $x^\star$ and $(u^\star, v^\star)$ are respectively primal and dual optimal LP solutions if and only if they solve:

$$\begin{aligned} A^T v + u &= c \\ x_i u_i &= 0, \; i = 1, \ldots, n \\ Ax &= b \\ x, u &\geq 0 \end{aligned}$$

Neat fact: the simplex method maintains the first three conditions and aims for the fourth one ... interior-point methods maintain the first and last two, and aim for the second

17.3. Primal-Dual Interior-Point Methods

The perturbed KKT conditions for standard form LP are hence:

$$
\begin{aligned}
A^T v + u &= c \\
x_i u_i &= 1/t, \; i = 1, \dots, n \\
Ax &= b \\
x, u &\geq 0
\end{aligned}
$$

What do our interior-point methods do?

Barrier (after eliminating u):

$$
\begin{aligned}
0 &= r_{\text{br}}(x, v) \\
&= \begin{pmatrix} A^T v + \operatorname{diag}(x)^{-1} \cdot (1/t)1 - c \\ Ax - b \end{pmatrix}
\end{aligned}
$$

Primal-dual:

$$
\begin{aligned}
0 &= r_{\text{pd}}(x, u, v) \\
&= \begin{pmatrix} A^T v + u - c \\ \operatorname{diag}(x)u - (1/t)1 \\ Ax - b \end{pmatrix}
\end{aligned}
$$

Barrier method: set $0 = r_{\rm br}(y + \Delta y) \approx r_{\rm br}(y) + Dr_{\rm br}(y)\Delta y$, i.e., solve

$$\begin{bmatrix} -\mathrm{diag}(x)^{-2}/t & A^T \\ A & 0 \end{bmatrix} \begin{pmatrix} \Delta x \\ \Delta v \end{pmatrix} = -r_{\rm br}(x, v)$$

and take a step $y^+ = y + s\Delta y$ (with line search for $s > 0$), and iterate until convergence. Then update $t = \mu t$

Primal-dual method: set $0 = r_{\rm pd}(y + \Delta y) \approx r_{\rm pd}(y) + Dr_{\rm pd}(y)\Delta y$, i.e., solve

$$\begin{bmatrix} 0 & I & A^T \\ \mathrm{diag}(u) & \mathrm{diag}(x) & 0 \\ A & 0 & 0 \end{bmatrix} \begin{pmatrix} \Delta x \\ \Delta u \\ \Delta v \end{pmatrix} = -r_{\rm pd}(x, u, v)$$

and take a step $y^+ = y + s\Delta y$ (with line search for $s > 0$), but only once. Then update $t = \mu t$

17.3. Primal-Dual Interior-Point Methods

Once backtracking allows for $s = 1$, i.e., we take one full Newton step, primal-dual method iterates will be primal and dual feasible from that point onwards

To see this, note that Δx, Δu, Δv are constructed so that

$$A^T \Delta v + \Delta u = -r_{\text{dual}} = -(A^T v + u - c)$$
$$A \Delta x = -r_{\text{prim}} = -(Ax - b)$$

Therefore after one full Newton step, $x^+ = x + \Delta x$, $u^+ = u + \Delta u$, $v^+ = v + \Delta v$, we have

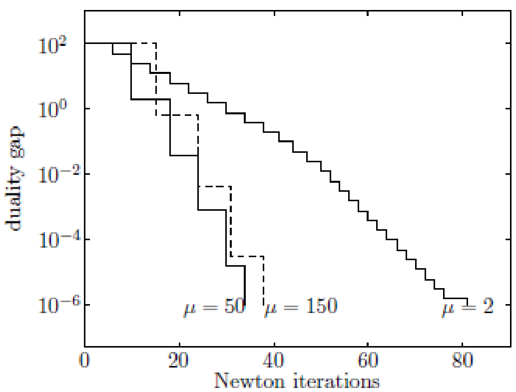
$$r^+_{\text{dual}} = A^T v^+ + u^+ - c = 0$$
$$r^+_{\text{prim}} = Ax^+ - b = 0,$$

so our iterates are primal and dual feasible

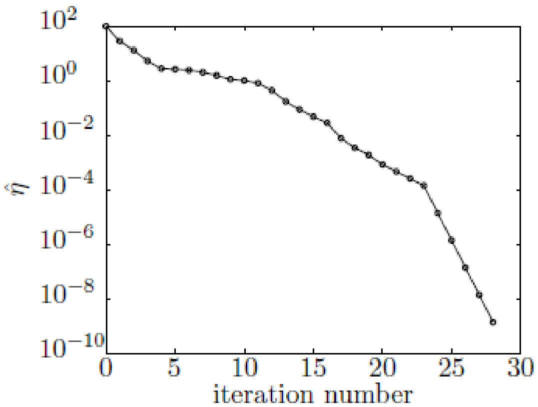
17.3. Primal-Dual Interior-Point Methods

Example from B & V 11.3.2 and 11.7.4: standard LP with $n = 50$ variables and $m = 100$ equality constraints

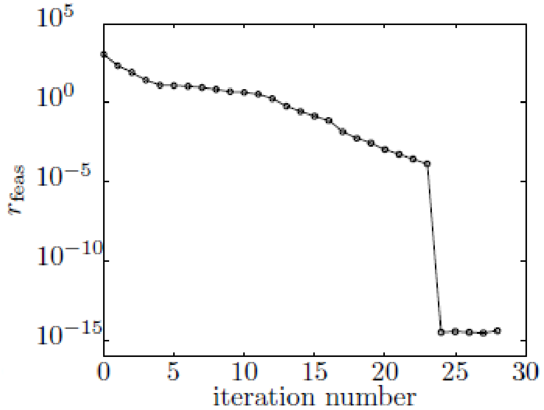
Barrier method uses various values of μ, primal-dual method uses $\mu = 10$. Both use $\alpha = 0.01$, $\beta = 0.5$



Barrier duality gap



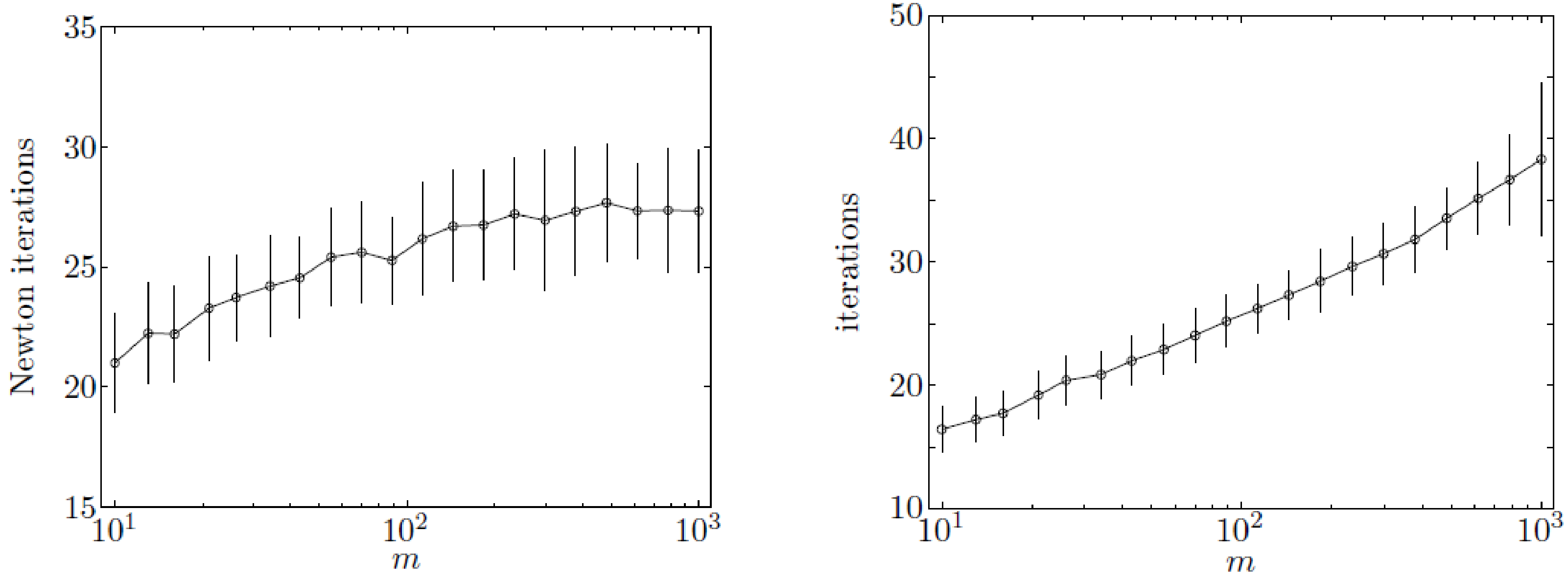
Primal-dual surrogate duality gap



Primal-dual feasibility gap, where $r_{\text{feas}} = (\|r_{\text{prim}}\|_2^2 + \|r_{\text{dual}}\|_2^2)^{1/2}$

Can see that primal-dual is faster to converge to high accuracy

Now a sequence of problems with $n = 2m$, and n growing. Barrier method uses $\mu = 100$, runs two outer loops (decreases duality gap by 10^4); primal-dual method uses $\mu = 10$, stops when surrogate duality gap and feasibility gap are at most 10^{-8}



Barrier method Primal-dual method

Primal-dual method requires only slightly more iterations, despite the fact that it is producing much higher accuracy solutions

17.3. Primal-Dual Interior-Point Methods

历史回顾

- Dantzig (1940s): the simplex method, still today is one of the most well-known/well-studied algorithms for LPs
- Klee and Minty (1972): pathological LP with n variables and $2n$ constraints, simplex method takes 2^n iterations to solve
- Khachiyan (1979): polynomial-time algorithm for LPs, based on ellipsoid method of Nemirovski and Yudin (1976). Strong in theory, weak in practice
- Karmarkar (1984): interior-point polynomial-time method for LPs. Fairly efficient (US Patent 4,744,026, expired in 2006)
- Renegar (1988): Newton-based interior-point algorithm for LP. Best known complexity ... until Lee and Sidford (2014)
- Modern state-of-the-art LP solvers typically use both simplex and interior-point methods

17.4. References

[1] S. Boyd, L. Vandenberghe, *Convex Optimization*, Cambridge University Press, 2004. http://www.stanford.edu/~boyd/cvxbook/ http://www.ee.ucla.edu/~vandenbe/cvxbook

[2] http://www.stat.cmu.edu/~ryantibs/convexopt/lectures/barr-method.pdf

[3] http://www.stat.cmu.edu/~ryantibs/convexopt/lectures/primal-dual.pdf