Convex Optimization Theory and Applications

Topic 16 - Equality Constrained Minimization

Li Li

Department of Automation
Tsinghua University

Fall, 2009-2021.

16.0. Outline

16.1. Equality Constraints

等式约束（凸）优化问题

$$\min\left\{ f(x) \middle| \text{ s.t. } Ax = b \right\}$$

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

基本假设：1） f 是具有连续二阶导数的凸函数

2） $A \in R^{p\times n}$, $\mathrm{rank}(A) = p < n$

3） $p^* = f(x^*) = \min\left\{ f(x) \middle| \text{ s.t. } Ax = b \right\}$ 为有限值

KKT 条件： $\exists x^* \in R^n, \nu^* \in R^p \;\Rightarrow\; \nabla f(x^*) + A^T \nu^* = 0,\ Ax^* = b$

16.1. Equality Constraints

此时我们一般有三种选择：

1. Eliminating equality constraints: $x = Fx + \hat{x}$, F spans null space of A, and $A\hat{x} = b$. Solve in terms of y

2. Deriving the dual: check that the Lagrange dual function is

$$g(v) = -b^T v + \inf_x \left(f(x) + v^T Ax \right) = -b^T v - \sup_x \left(-f(x) - \left(A^T v \right)^T x \right) = -f^*(A^T v) - b^T v$$

$f^*(\cdot)$ is the conjugate of $f(\cdot)$, and strong duality holds. With luck, we can express optimal primal variable x^* in terms of optimal dual variable v^*

3. Equality-constrained Newton: in many cases, this is the most straightforward option

16.1. Equality Constraints

选择一：

假设已知 $\hat{x} \in \operatorname{dom} f$ 满足 $A\hat{x} = b$

用 $F \in R^{n\times(n-p)}$ 表示 A 的零空间 $\{x \mid Ax = 0\}$ 的基矩阵（$AF = 0$）

$$\operatorname{rank}(F) = n - p\,, \quad \{x \mid Ax = 0\} = \left\{Fz \mid z \in R^{n-p}\right\}$$

$$\Rightarrow \qquad \{x \mid Ax = b\} = \left\{Fz + \hat{x} \mid z \in R^{n-p}\right\}$$

$$\min\{f(x) \mid \text{s.t. } Ax = b\} \quad \Leftrightarrow \quad \min_{z \in R^{n-p}} \tilde{f}(z) = f(Fz + \hat{x})$$

16.1. Equality Constraints

example: optimal allocation with resource constraint

$$\begin{array}{ll} \text{minimize} & f_1(x_1) + f_2(x_2) + \cdots + f_n(x_n) \\ \text{subject to} & x_1 + x_2 + \cdots + x_n = b \end{array}$$

eliminate $x_n = b - x_1 - \cdots - x_{n-1}$, *i.e.*, choose

$$\hat{x} = be_n, \qquad F = \begin{bmatrix} I \\ -\mathbf{1}^T \end{bmatrix} \in \mathbf{R}^{n \times (n-1)}$$

reduced problem:

$$\text{minimize} \quad f_1(x_1) + \cdots + f_{n-1}(x_{n-1}) + f_n(b - x_1 - \cdots - x_{n-1})$$

(variables $x_1, \ldots, x_{n-1}$)

16.1. Equality Constraints

选择二：

例子、等式约束的凸二次规划

$$\min\left\{ f(x) = \frac{1}{2}x^T P x + q^T x + r \,\middle|\, \text{s.t. } Ax = b \right\},\quad P \in S_+^n,\quad A\text{行满秩}$$

根据 KKT 方程$\begin{bmatrix} P & A^T \\ A & 0 \end{bmatrix}\begin{bmatrix} x \\ \nu \end{bmatrix} = \begin{bmatrix} -q \\ b \end{bmatrix}$解的情况可确定三种可能：

1）唯一最优解；2）无穷多最优解；3）无下界

可证：在给定条件下，$\begin{bmatrix} P & A^T \\ A & 0 \end{bmatrix}$非奇异等价于$\begin{bmatrix} P \\ A \end{bmatrix}$列满秩

或者等价的$P + A^T A > 0$

Example 10.2 *Equality constrained analytic center.* We consider the problem

$$\begin{array}{ll} \text{minimize} & f(x) = -\sum_{i=1}^n \log x_i \\ \text{subject to} & Ax = b, \end{array} \tag{10.7}$$

where $A \in \mathbf{R}^{p \times n}$, with implicit constraint $x \succ 0$. Using

$$f^*(y) = \sum_{i=1}^n (-1 - \log(-y_i)) = -n - \sum_{i=1}^n \log(-y_i)$$

(with $\mathbf{dom}\, f^* = -\mathbf{R}^n_{++}$), the dual problem is

$$\text{maximize} \quad g(\nu) = -b^T\nu + n + \sum_{i=1}^n \log(A^T\nu)_i, \tag{10.8}$$

with implicit constraint $A^T\nu \succ 0$. Here we can easily solve the dual feasibility equation, *i.e.*, find the x that minimizes $L(x, \nu)$:

$$\nabla f(x) + A^T\nu = -(1/x_1, \ldots, 1/x_n) + A^T\nu = 0,$$

and so

$$x_i(\nu) = 1/(A^T\nu)_i. \tag{10.9}$$

To solve the equality constrained analytic centering problem (10.7), we solve the (unconstrained) dual problem (10.8), and then recover the optimal solution of (10.7) via (10.9).

选择三：

In equality-constrained Newton's method, we start with $x^{(0)}$ such that $Ax^{(0)} = b$. Then we repeat the updates

$$x^+ = x + tv, \quad \text{where}$$

$$v = \underset{Az=0}{\operatorname{argmin}} \; \nabla f(x)^T(z-x) + \frac{1}{2}(z-x)^T \nabla^2 f(x)(z-x)$$

This keeps x^+ in feasible set, since $Ax^+ = Ax + tAv = b + 0 = b$

Furthermore, v is the solution to minimizing a quadratic subject to equality constraints. We know from KKT conditions that v satisfies

$$\begin{bmatrix} \nabla^2 f(x) & A^T \\ A & 0 \end{bmatrix} \begin{bmatrix} v \\ w \end{bmatrix} = \begin{bmatrix} -\nabla f(x) \\ 0 \end{bmatrix}$$

for some w. Hence Newton direction v is again given by solving a linear system in the Hessian (albeit a bigger one)

16.2. Newton's Method with Eq. Constraints

推导途径：用 Newton 方法解无约束问题 $\min_{z\in R^{n-p}} \tilde{f}(z)=f(Fz+\hat{x})$

偏导数： $\nabla\tilde{f}(z)=F^T\nabla f(Fz+\hat{x})$， $\nabla^2\tilde{f}(z)=F^T\nabla^2 f(Fz+\hat{x})F$

牛顿方向： $d_z=-\left(F^T\nabla^2 f(Fz+\hat{x})F\right)^{-1}\left(F^T\nabla f(Fz+\hat{x})\right)$

直线搜索： $z'=z+td_z$， $\tilde{f}(z')\le\tilde{f}(z)+\alpha t\nabla\tilde{f}(z)^T d_z$

停止准则： $\frac{1}{2}d_z^T\nabla^2\tilde{f}(z)d_z=\frac{1}{2}\left|\nabla\tilde{f}(z)^T d_z\right|\le\varepsilon$

16.2. Newton's Method with Eq. Constraints

记 $x' = Fz' + \hat{x},\ x = Fz + \hat{x},\ d_x = Fd_z$ （$\nabla\tilde{f}(z)^T d_z = \nabla f(x)^T d_x$）

确定 d_x 的方程： $d_z = -\left(F^T \nabla^2 f(Fz+\hat{x})F\right)^{-1}\left(F^T \nabla f(Fz+\hat{x})\right)$

$$\Leftrightarrow \quad F^T\left(\nabla^2 f(x)d_x + \nabla f(x)\right) = 0 \quad \Leftrightarrow \quad \nabla^2 f(x)d_x + \nabla f(x) = A^T \nu$$

$$d_x = Fd_z \quad \Rightarrow \quad Ad_x = 0$$

直线搜索规则： $z' = z + td_z,\ \ \tilde{f}(z') \le \tilde{f}(z) + \alpha t \nabla\tilde{f}(z)^T d_z$

$$\Leftrightarrow \quad x' = x + td_x,\ \ f(x') \le f(x) + \alpha t \nabla f(x)^T d_x$$

停止准则： $\frac{1}{2}\left|\nabla\tilde{f}(z)^T d_z\right| = \frac{1}{2}\left|\nabla f(x)^T d_x\right| \le \varepsilon$

16.2. Newton's Method with Eq. Constraints

总结：在已知 $Ax=b$ 的一个解的情况下，将等式约束变成无约束优化问题，并用 Newton 方法求解，等价于解下述线性方程组（x 给定）确定搜索方向，然后对原函数直线搜索

$$\begin{pmatrix} \nabla^2 f(x) & A^T \\ A & 0 \end{pmatrix} \begin{pmatrix} d_x \\ \nu \end{pmatrix} = \begin{pmatrix} -\nabla_x f(x) \\ 0 \end{pmatrix}$$

停止准则同样可写成：$\frac{1}{2} d_x{}^T \nabla^2 f(x) d_x \le \varepsilon$

理由：$\nabla^2 f(x) d_x + \nabla f(x) = A^T \nu,\ A d_x = 0 \quad \Rightarrow \quad \left|\nabla f(x)^T d_x\right| = d_x{}^T \nabla^2 f(x) d_x$

16.2. Newton's Method with Eq. Constraints

等式约束 Newton 方法的其它解释

1）对目标函数二阶近似

$$f(x+d) \approx f(x) + \nabla f(x)^T d + \frac{1}{2} d^T \nabla^2 f(x) d \triangleq \bar{f}_x(d)$$

$$\min\left\{ f\left(x+d\right) \middle| \text{ s.t. } A\left(x+d\right)=b \right\} \approx \min\left\{ \bar{f}_x(d) \middle| \text{ s.t. } Ad=0 \right\}$$

$$\Rightarrow \quad \begin{pmatrix} \nabla^2 f(x) & A^T \\ A & 0 \end{pmatrix} \begin{pmatrix} d \\ \nu \end{pmatrix} = \begin{pmatrix} -\nabla f(x) \\ 0 \end{pmatrix}$$

16.2. Newton's Method with Eq. Constraints

2）对目标函数的梯度进行一阶近似

给定 x，求解等式约束问题等价于求解非线性 KKT 方程

$$\nabla f(x+d)+A^T\nu=0,\ \ A(x+d)=b$$

$$\nabla f(x+d)\approx\nabla f(x)+\nabla^2 f(x)d$$

$$\Rightarrow \qquad \begin{pmatrix}\nabla^2 f(x) & A^T\\ A & 0\end{pmatrix}\begin{pmatrix}d\\ \nu\end{pmatrix}=\begin{pmatrix}-\nabla f(x)\\ 0\end{pmatrix}$$

16.2. Newton's Method with Eq. Constraints

收敛性分析假设

$\tilde{f}(z) = f(Fz + \hat{x})$ 要满足无约束问题牛顿方法的假设：

$$\tilde{S} = \left\{ \mathrm{z} \in \mathrm{dom}\, \tilde{f} \,\middle|\, \tilde{f}(z) \le \tilde{f}\left(z^{(0)}\right) \right\}$$ 是闭集

$$mI \le \nabla^2 \tilde{f}(z) \le MI,\ \ \forall \mathrm{z} \in \tilde{S} \triangleq \left\{ \mathrm{z} \in \mathrm{dom}\, \tilde{f} \,\middle|\, \tilde{f}(z) \le \tilde{f}\left(z^{(0)}\right) \right\}$$

$$\left\| \nabla^2 \tilde{f}(y) - \nabla^2 \tilde{f}(z) \right\| \le \tilde{L} \| y - z \|, \forall y, z \in \tilde{S}$$

16.2. Newton's Method with Eq. Constraints

能满足以上要求的假设

1） $S=\left\{x \middle| x\in \operatorname{dom} f,\ f\left(x\right)\le f\left(x^0\right),\ Ax=b\right\}$是闭集，其中

$x^0 \in \operatorname{dom} f,\ Ax^0=b$

2） $\nabla^2 f\left(x\right)\le MI,\ \forall x\in S$ 以及

$$\left\|\begin{pmatrix}\nabla^2 f(x) & A^T\\ A & 0\end{pmatrix}^{-1}\right\|\le K$$

3） $\left\|\nabla^2 f\left(x\right)-\nabla^2 f\left(y\right)\right\|\le L\left\|x-y\right\|,\ \forall x,y\in S$

16.2. Newton's Method with Eq. Constraints

第二条假设的作用

对任意 $z \in \tilde{S}$，记 $x = \hat{x} + Fz$，用 u 表示 $\nabla^2 \tilde{f}(z) = F^T \nabla^2 f(x) F$ 的最小特征根对应的单位化特征向量，即 $\nabla^2 \tilde{f}(z) u = \lambda_{\min}\left(\nabla^2 \tilde{f}(z)\right) u$，$\|u\| = 1$。因为 $AF = 0$

$$\Rightarrow \qquad \begin{pmatrix} \nabla^2 f(x) & A^T \\ A & 0 \end{pmatrix} \begin{pmatrix} Fu \\ 0 \end{pmatrix} = \begin{pmatrix} \nabla^2 f(x) Fu \\ 0 \end{pmatrix}$$

$$\Rightarrow \qquad \begin{pmatrix} \nabla^2 f(x) & A^T \\ A & 0 \end{pmatrix}^{-1} \begin{pmatrix} \nabla^2 f(x) Fu \\ 0 \end{pmatrix} = \begin{pmatrix} Fu \\ 0 \end{pmatrix}$$

16.2. Newton's Method with Eq. Constraints

$$\lambda_{\min}\left(F^T F\right)^{\frac{1}{2}} \le \left(u^T F^T F u\right)^{\frac{1}{2}} = \left\|Fu\right\| \le \left\|\begin{pmatrix} \nabla^2 f(x) & A^T \\ A & 0 \end{pmatrix}^{-1}\right\| \left\|\nabla^2 f(x) F u\right\|$$

$$\Rightarrow \quad \le K \left\|\nabla^2 f(x)^{\frac{1}{2}}\right\| \left\|\nabla^2 f(x)^{\frac{1}{2}} F u\right\|$$

$$\le K M^{\frac{1}{2}} \left(u^T \nabla^2 \tilde{f}(z) u\right)^{\frac{1}{2}}$$

$$= K M^{\frac{1}{2}} \lambda_{\min}\left(\nabla^2 \tilde{f}(z)\right)^{\frac{1}{2}}$$

$$\Rightarrow \quad \lambda_{\min}\left(\nabla^2 \tilde{f}(z)\right) \ge \frac{\lambda_{\min}\left(F^T F\right)}{K^2 M}$$

16.2. Newton's Method with Eq. Constraints

Infeasible start Newton method

---

given starting point $x \in \mathbf{dom}\, f$, ν, tolerance $\epsilon > 0$, $\alpha \in (0, 1/2)$, $\beta \in (0, 1)$.

repeat

1. Compute primal and dual Newton steps Δx_{nt}, $\Delta \nu_{\mathrm{nt}}$.
2. *Backtracking line search on* $\|r\|_2$.

   $t := 1$.

   while $\|r(x + t\Delta x_{\mathrm{nt}}, \nu + t\Delta \nu_{\mathrm{nt}})\|_2 > (1 - \alpha t)\|r(x, \nu)\|_2$, $\quad t := \beta t$.
3. *Update.* $x := x + t\Delta x_{\mathrm{nt}}$, $\nu := \nu + t\Delta \nu_{\mathrm{nt}}$.

until $Ax = b$ and $\|r(x, \nu)\|_2 \leq \epsilon$.

- not a descent method: $f(x^{(k+1)}) > f(x^{(k)})$ is possible
- directional derivative of $\|r(y)\|_2^2$ in direction $\Delta y = (\Delta x_{\mathrm{nt}}, \Delta \nu_{\mathrm{nt}})$ is

$$\frac{d}{dt} \|r(y + \Delta y)\|_2 \bigg|_{t=0} = -\|r(y)\|_2$$

16.2. Newton's Method with Eq. Constraints

不可行初始点 Newton 方法

基本想法：从任意 $x^0 \in \operatorname{dom} f$，$\nu^0 \in R^p$ 开始求解 KKT 方程

$$\nabla f(x) + A^T \nu = 0, \quad Ax = b$$

记 $y = \begin{pmatrix} x \\ \nu \end{pmatrix}$，定义原对偶残差 $r(y) = \begin{pmatrix} \nabla f(x) + A^T \nu) \\ Ax - b \end{pmatrix} \triangleq \begin{pmatrix} r_{\text{dual}}(y) \\ r_{\text{pri}}(y) \end{pmatrix}$

$\Rightarrow$ 求 $d_y = \begin{pmatrix} d_x \\ d_\nu \end{pmatrix}$ 满足 $r(y + d_y) = 0 \quad \Rightarrow \quad \min_{d_y} \left\| r(y + d_y) \right\|$

16.2. Newton's Method with Eq. Constraints

对残差向量进行一阶近似

$$r\left(y+d_y\right) \approx r\left(y\right)+Dr\left(y\right)d_y$$

其中

$$Dr\left(y\right)=\frac{\partial r\left(y\right)}{\partial y^T}=\begin{pmatrix} \nabla^2 f(x) & A^T \\ A & 0 \end{pmatrix}$$

$$r\left(y\right)+Dr\left(y\right)d_y=0$$

$\Rightarrow$ 不可行初始点牛顿方向：

$$d_y=-Dr\left(y\right)^{-1}r\left(y\right)=-\begin{pmatrix} \nabla^2 f(x) & A^T \\ A & 0 \end{pmatrix}^{-1}\begin{pmatrix} \nabla f(x)+A^T\nu \\ Ax-b \end{pmatrix} \triangleq \begin{pmatrix} d_x \\ d_\nu \end{pmatrix}$$

16.2. Newton's Method with Eq. Constraints

回溯直线搜索

定义 $g(t) = \left\| r\left(y + td_y\right) \right\| \quad \Rightarrow \quad \left(g(t)\right)^2 = r\left(y + td_y\right)^T r\left(y + td_y\right)$

$$\Rightarrow \quad g(t)g'(t) = r\left(y + td_y\right)^T Dr\left(y + td_y\right) d_y$$

$$\Rightarrow \quad g'(0) = -\left\| r\left(y\right) \right\|$$

直线搜索准则 $g(t) \le g(0) + \alpha t g'(0)$

$$\Leftrightarrow \quad \left\| r\left(y + td_y\right) \right\| \le \left(1 - \alpha t\right) \left\| r\left(y\right) \right\|$$

16.2. Newton's Method with Eq. Constraints

原残差和步长的关系

$$\begin{pmatrix} \nabla^2 f(x) & A^T \\ A & 0 \end{pmatrix}\begin{pmatrix} d_x \\ d_\nu \end{pmatrix} = -\begin{pmatrix} \nabla f(x) + A^T \nu \\ Ax - b \end{pmatrix} \quad \Rightarrow \quad Ad_x = -(Ax - b)$$

$$\Rightarrow \quad r_{\text{pri}}(y + td_y) = A(x + td_x) - b = (1-t)(Ax - b) = (1-t)\, r_{\text{pri}}(y)$$

$$\Rightarrow \quad r_{\text{pri}}(y^k) = \left(\prod_{i=0}^{k-1}(1-t^i)\right) r_{\text{pri}}(y^0) \quad \Rightarrow \quad t^{\hat{k}} = 1\text{，}\ \text{成立}\ Ax^k = b,\ \forall k \ge \hat{k}$$

$\Rightarrow$ 算法停止准则：$Ax^k = b$ 以及 $\left\| r_{\text{dual}}(y^k) \right\| \le \varepsilon$

16.2. Newton's Method with Eq. Constraints

用于收敛性分析的关键不等式

$$\left\| r\left(y+td_y\right)\right\| \le (1-t)\left\| r\left(y\right)\right\| + \frac{1}{2}K^2Lt^2\left\| r\left(y\right)\right\|^2 \triangleq \varphi\left(t\right)$$

推导该不等式的条件

$$r\left(y\right) = -Dr\left(y\right)d_y$$

$$r\left(y+td_y\right) = r\left(y\right) + \int_0^1 Dr\left(y+\tau td_y\right)td_y d\tau$$

$$\left\| Dr\left(y\right)^{-1}\right\| \le K$$

$$\left\| \nabla^2 f\left(x\right) - \nabla^2 f\left(z\right)\right\| \le L\left\| x-z\right\|$$

16.2. Newton's Method with Eq. Constraints

收敛性分析（确定满足给定误差阈值的迭代次数上界）

确定η满足$\left\|r\left(y^k\right)\right\| \le \eta \;\Rightarrow\; t^k = 1,\; \left\|r\left(y^{k+1}\right)\right\| \le \eta \;\Rightarrow$ 二次收敛

利用直线搜索停止准则 $\left\|r\left(y^k + t^k d_{y^k}\right)\right\| \le \left(1-\alpha t^k\right)\left\|r\left(y^k\right)\right\|$和不等式 $\left\|r\left(y^k + d_{y^k}\right)\right\| \le \varphi(1) = \frac{1}{2}K^2 L \left\|r\left(y^k\right)\right\|^2$

若 $\left\|r\left(y^{(k)}\right)\right\| \le \eta \le \dfrac{2(1-\alpha)}{K^2 L} \quad \Rightarrow \quad \left\|r\left(y^k + d_{y^k}\right)\right\| \le \varphi(1) \le (1-\alpha)\left\|r\left(y^k\right)\right\|$

$$\Rightarrow \quad t^k = 1,\quad \left\|r\left(y^{k+1}\right)\right\| \le \eta$$

16.2. Newton's Method with Eq. Constraints

此时可得 $$\frac{1}{2}K^2L\left\|r\left(y^{k+1}\right)\right\| \le \frac{1}{2}K^2L\varphi(1) \le \left(\frac{1}{2}K^2L\left\|r\left(y^k\right)\right\|\right)^2$$

于是 $$\frac{1}{2}K^2L\left\|r\left(y^{k+\tau}\right)\right\| \le \left(\frac{1}{2}K^2L\left\|r\left(y^k\right)\right\|\right)^{2^\tau}, \quad \forall \tau \ge 1$$

如果进一步要求$\left\|r\left(y^k\right)\right\| \le \eta \le \frac{1}{K^2L}$ $\left(\le \frac{2(1-\alpha)}{K^2L}\right)$, 又可得到

$$\frac{1}{2}K^2L\left\|r\left(y^{k+\tau}\right)\right\| \le \left(\frac{1}{2}\right)^{2^\tau}, \quad \forall \tau \ge 1$$

16.2. Newton's Method with Eq. Constraints

二次收敛阶段迭代次数上界

如果 k 是首次满足 $\left\|r\left(y^k\right)\right\| \le \eta \triangleq \dfrac{1}{K^2 L}$ 的迭代次数, 则有

$$\left\|r\left(y^{k+\tau}\right)\right\| \le \varepsilon_0 \left(\frac{1}{2}\right)^{2^\tau} \quad (\text{其中 } \varepsilon_0 = \frac{2}{K^2 L})$$

至多再迭代 $K_2 \ge \log_2\left(\log_2\left(\dfrac{\varepsilon_0}{\varepsilon}\right)\right)$ 次

就可满足 $\left\|r\left(y^{k+K_2}\right)\right\| \le \varepsilon$

16.2. Newton's Method with Eq. Constraints

阻尼 Newton 阶段迭代次数上界

对任意的$\left\|r\left(y^k\right)\right\| > \eta = \dfrac{1}{K^2 L}$，令 $\bar{t} = \dfrac{1}{K^2 L\left\|r\left(y^k\right)\right\|}$，则有

$$\bar{t} < 1,\quad \left\|r\left(y^k + \bar{t} d_{y^k}\right)\right\| \le \varphi(\bar{t}) = \left(1 - \frac{1}{2}\bar{t}\right)\left\|r\left(y^k\right)\right\| \le (1 - \alpha\bar{t})\left\|r\left(y^k\right)\right\|$$

$$\Rightarrow \quad \beta^{-1} t^k > \bar{t} \quad \Rightarrow \quad t^k > \beta\bar{t}$$

$$\Rightarrow \quad \left\|r\left(y^k + t^k d_{y^k}\right)\right\| \le \left(1 - \alpha t^k\right)\left\|r\left(y^k\right)\right\| \le (1 - \alpha\beta\bar{t})\left\|r\left(y^k\right)\right\| = \left\|r\left(y^k\right)\right\| - \alpha\beta\eta$$

$$\Rightarrow \quad \left\|r\left(y^k\right)\right\| \le \left\|r\left(y^0\right)\right\| - k\alpha\beta\eta$$

$$\eta < \left\|r\left(y^{K_1}\right)\right\| \le \left\|r\left(y^0\right)\right\| - K_1\alpha\beta\eta \quad \Rightarrow \quad K_1 \le \frac{\left\|r\left(y^0\right)\right\| - \eta}{\alpha\beta\eta}$$

16.2. Newton's Method with Eq. Constraints

最后说明原对偶点列的收敛性

对任意的 $t \geq k$ 和 $l \geq 1$

$$\left\| y^{t+l} - y^t \right\| \leq \sum_{i=0}^{l-1} \left\| y^{t+i+1} - y^{t+i} \right\| = \sum_{i=0}^{l-1} \left\| Dr\left(y^{t+i}\right)^{-1} r\left(y^{t+i}\right) \right\|$$

$$\leq K \sum_{i=0}^{l-1} \left\| r\left(y^{t+i}\right) \right\| \leq K\varepsilon_0 \sum_{i=0}^{l-1} \left(\frac{1}{2}\right)^{2^{t+i-k}}$$

$$\leq K\varepsilon_0 \left(\frac{1}{2}\right)^{2^{t-k}} \sum_{i=0}^{l-1} \left(\frac{1}{2}\right)^{2^i} \leq \left(\frac{1}{2}\right)^{2^{t-k}} \hat{C}$$

$$\Rightarrow \quad \lim_{k \to \infty} y^k = y^*, \quad \left\| r\left(y^*\right) \right\| = 0$$

16.2. Newton's Method with Eq. Constraints

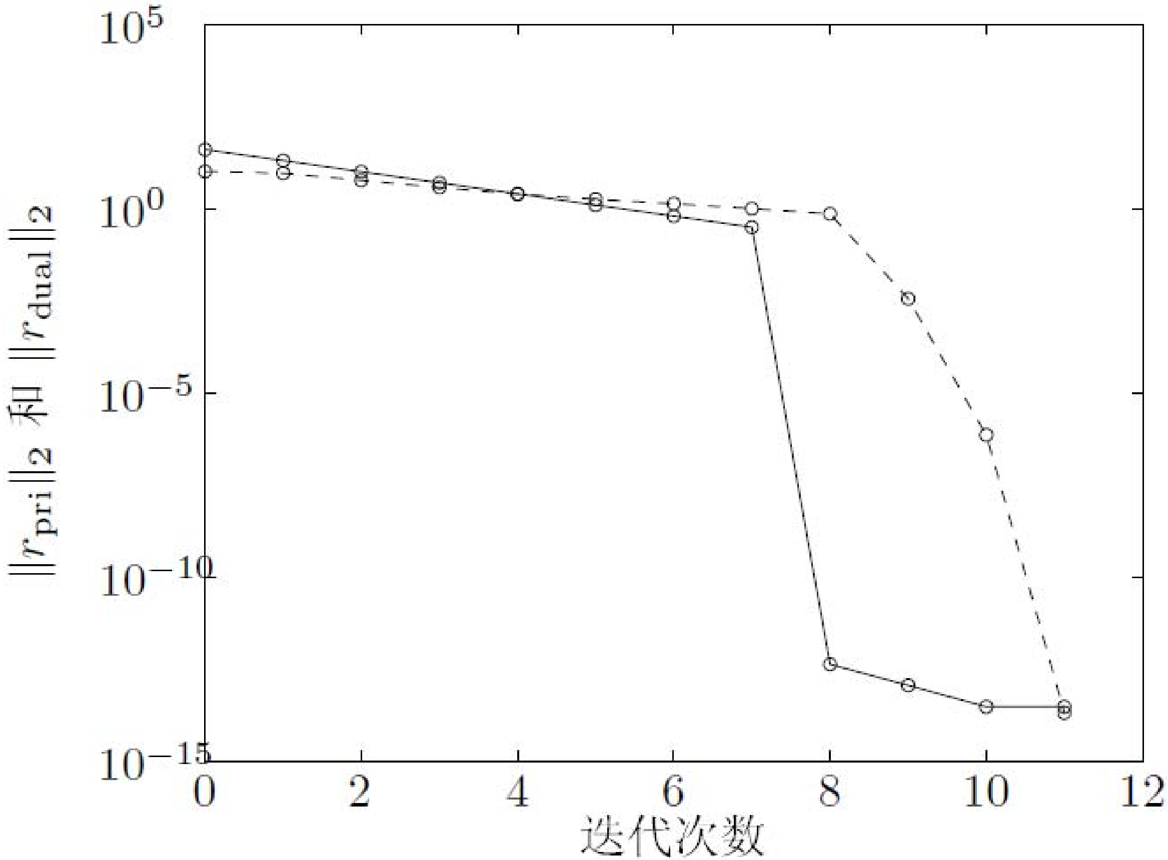


图 10.1 用不可行初始点 Newton 方法求解一个 100 个变量，50 个等式约束的解析中心点问题。图中给出了 $\|r_{\mathrm{pri}}\|_2$（实线）和 $\|r_{\mathrm{dual}}\|_2$（虚线）。可行性在 8 次迭代后满足（并一直保持），大约 9 次迭代后开始二次收敛。

16.2. Newton's Method with Eq. Constraints

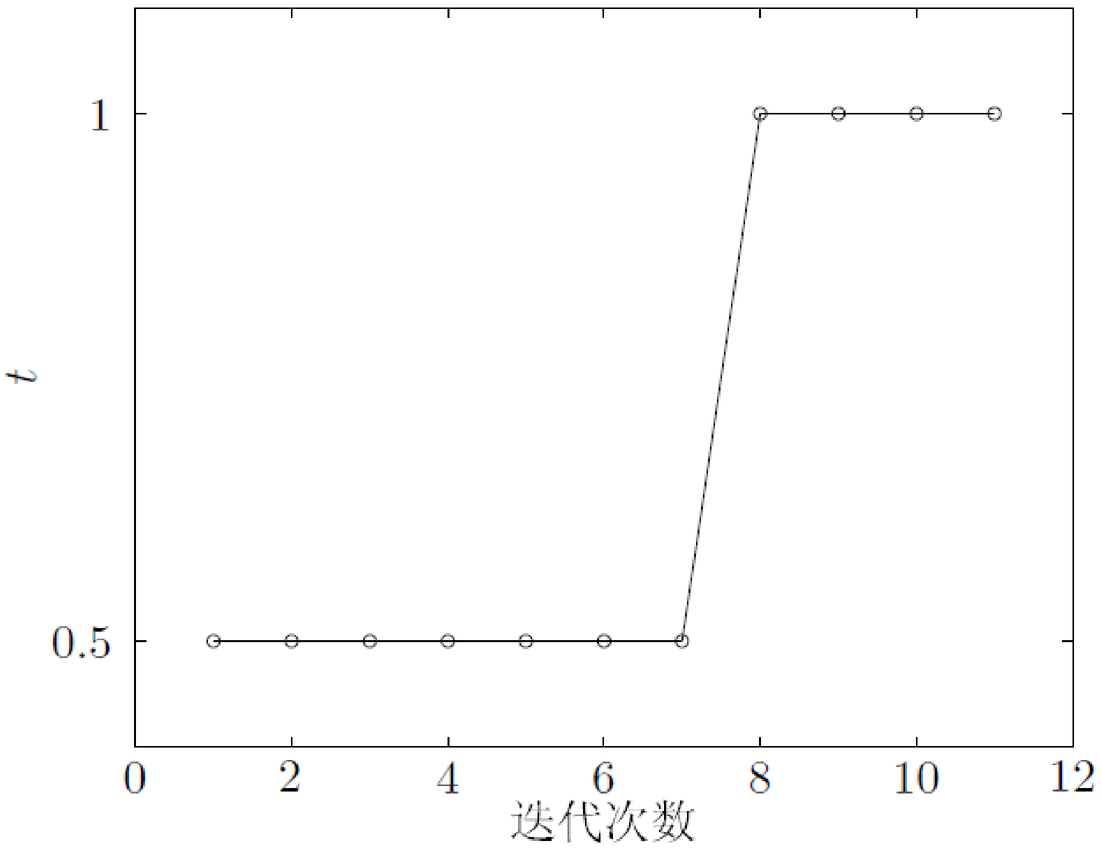


图 10.2 相同例子的步长与迭代次数之间的关系。第 8 次迭代时选取了完整步长，自此以后始终保持了可行性。

16.2. Newton's Method with Eq. Constraints

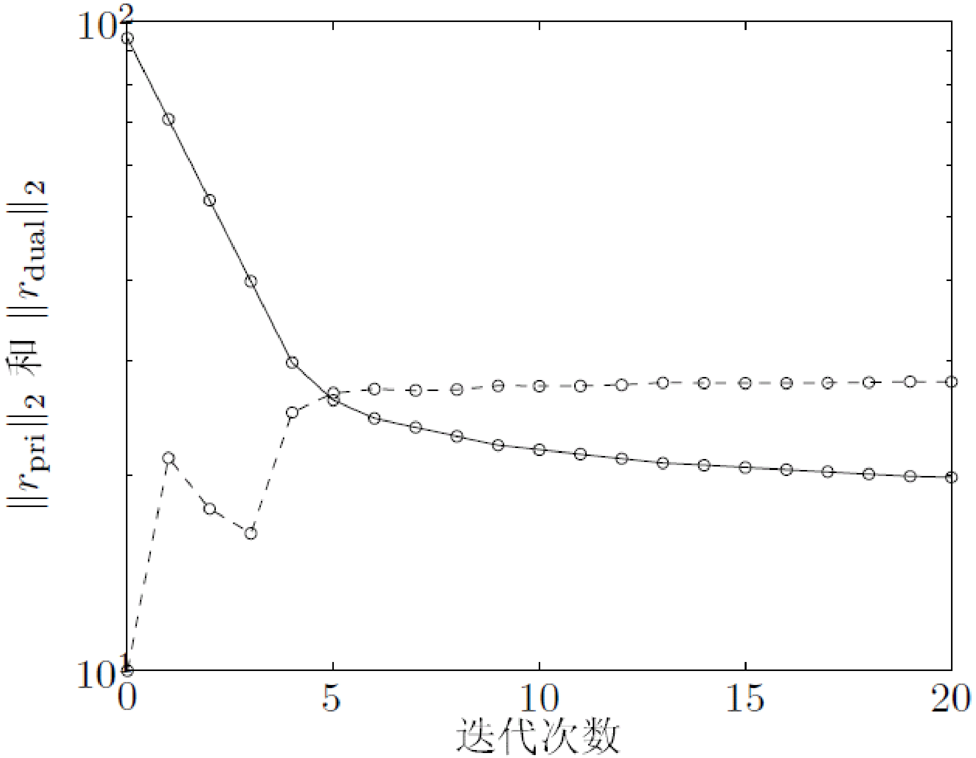


图 10.3 用不可行初始点 Newton 方法求解一个 100 个变量，50 个等式约束的解析中心点问题，其定义域 $\mathbf{dom}\, f = \mathbf{R}^{100}_{++}$ 和 $\{z \mid Az = b\}$ 不相交。图中给出了 $\|r_{\mathrm{pri}}\|_2$（实线）和 $\|r_{\mathrm{dual}}\|_2$（虚线）。在这种情况下，残差不收敛于 0。

16.2. Newton's Method with Eq. Constraints

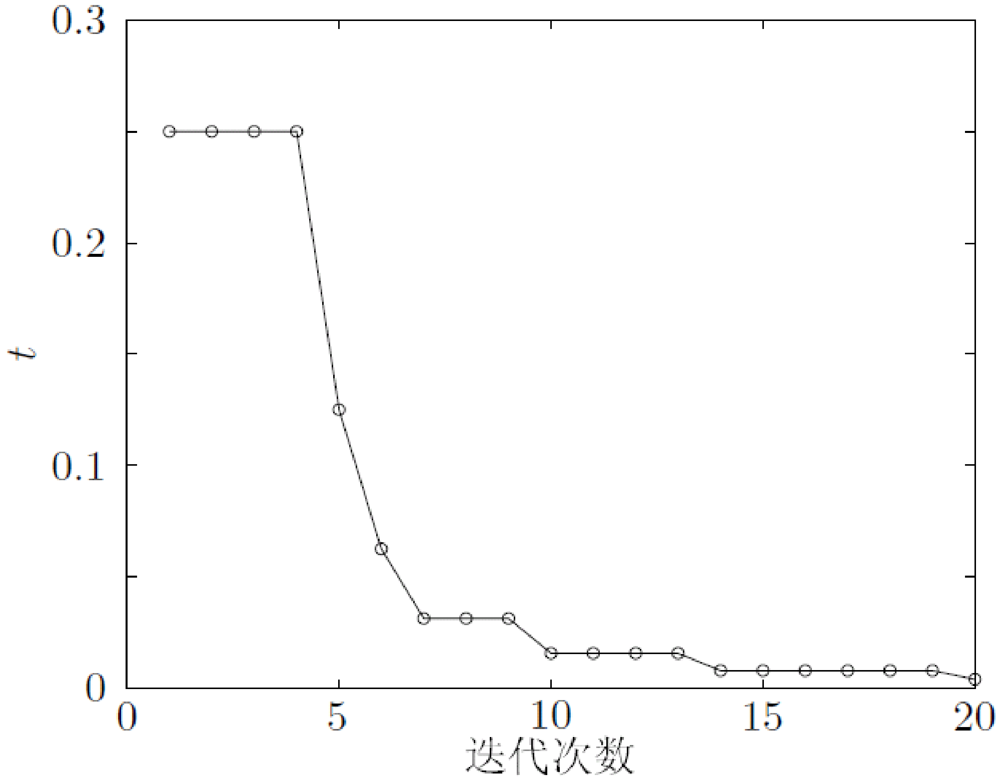


图 10.4 不可行实例的步长和迭代次数之间的关系。步长永远不等于 1，并趋近于 0。

16.2. Newton's Method with Eq. Constraints

主要计算量是解下述类型的方程组确定 Newton 方向

$$Hv + A^T w = -g, \quad Av = -h$$

消除 v 可解得 $w = \left(AH^{-1}A^T\right)^{-1}\left(h - AH^{-1}g\right)$

然后算出 $v = -H^{-1}\left(g + A^T w\right)$

其中可利用 Cholesky 分解和前后向代入节约计算量

16.2. Newton's Method with Eq. Constraints

Network flow optimization

$$\begin{array}{ll} \text{minimize} & \sum_{i=1}^{n} \phi_i(x_i) \\ \text{subject to} & Ax = b \end{array}$$

- directed graph with n arcs, $p+1$ nodes
- x_i: flow through arc i; ϕ_i: cost flow function for arc i (with $\phi_i''(x) > 0$)
- node-incidence matrix $\tilde{A} \in \mathbf{R}^{(p+1)\times n}$ defined as

$$\tilde{A}_{ij} = \begin{cases} 1 & \text{arc } j \text{ leaves node } i \\ -1 & \text{arc } j \text{ enters node } i \\ 0 & \text{otherwise} \end{cases}$$

- reduced node-incidence matrix $A \in \mathbf{R}^{p\times n}$ is $\tilde{A}$ with last row removed
- $b \in \mathbf{R}^p$ is (reduced) source vector
- $\mathbf{rank}\, A = p$ if graph is connected

16.2. Newton's Method with Eq. Constraints

KKT system

$$\begin{bmatrix} H & A^T \\ A & 0 \end{bmatrix} \begin{bmatrix} v \\ w \end{bmatrix} = - \begin{bmatrix} g \\ h \end{bmatrix}$$

- $H = \mathbf{diag}(\phi_1''(x_1), \ldots, \phi_n''(x_n))$, positive diagonal
- solve via elimination:

$$AH^{-1}A^T w = h - AH^{-1}g, \qquad Hv = -(g + A^T w)$$

sparsity pattern of coefficient matrix is given by graph connectivity

$$\begin{aligned} (AH^{-1}A^T)_{ij} \neq 0 \quad & \Longleftrightarrow \quad (AA^T)_{ij} \neq 0 \\ & \Longleftrightarrow \quad \text{nodes } i \text{ and } j \text{ are connected by an arc} \end{aligned}$$

16.3. Analytic Center of LMI

the analytic center of set of convex inequalities and linear equations

$$f_i(x) \le 0, \quad i = 1, \dots, m, \qquad Fx = g$$

is defined as the optimal point of

$$\begin{array}{ll} \text{minimize} & -\sum_{i=1}^{m} \log(-f_i(x)) \\ \text{subject to} & Fx = g \end{array}$$

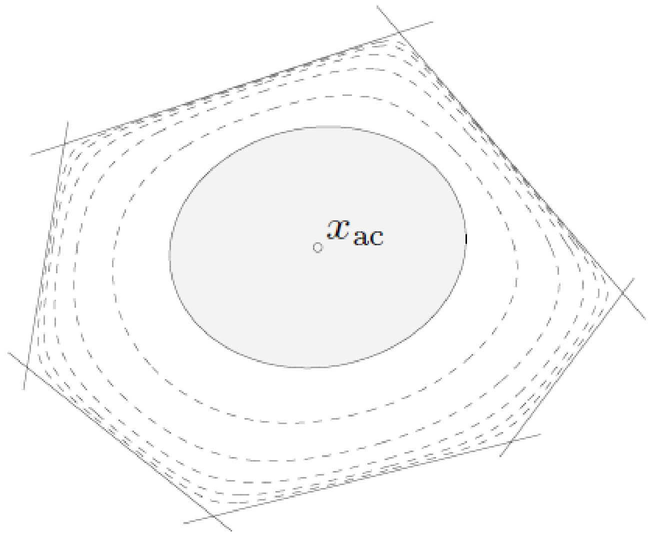
- more easily computed than MVE or Chebyshev center (see later)
- not just a property of the feasible set: two sets of inequalities can describe the same set, but have different analytic centers

16.3. Analytic Center of LMI

analytic center of linear inequalities $a_i^T x \le b_i,\ i = 1, \dots, m$

x_{ac} is minimizer of

$$\phi(x) = -\sum_{i=1}^{m} \log(b_i - a_i^T x)$$



inner and outer ellipsoids from analytic center:

$$\mathcal{E}_{\mathrm{inner}} \subseteq \{x \mid a_i^T x \le b_i,\ i = 1, \dots, m\} \subseteq \mathcal{E}_{\mathrm{outer}}$$

where

$$\begin{aligned}
\mathcal{E}_{\mathrm{inner}} &= \{x \mid (x - x_{\mathrm{ac}})^T \nabla^2 \phi(x_{\mathrm{ac}})(x - x_{\mathrm{ac}} \le 1\} \\
\mathcal{E}_{\mathrm{outer}} &= \{x \mid (x - x_{\mathrm{ac}})^T \nabla^2 \phi(x_{\mathrm{ac}})(x - x_{\mathrm{ac}}) \le m(m-1)\}
\end{aligned}$$

16.3. Analytic Center of LMI

Newton's method for analytic center of LMI

$$\begin{array}{ll} \text{minimize} & -\log\det X \\ \text{subject to} & \mathbf{tr}(A_i X) = b_i, \quad i = 1, \ldots, p \end{array}$$

variable $X \in \mathbf{S}^n$

optimality conditions

$$X^\star \succ 0, \qquad -(X^\star)^{-1} + \sum_{j=1}^{p} \nu_j^\star A_i = 0, \qquad \mathbf{tr}(A_i X^\star) = b_i, \quad i = 1, \ldots, p$$

Newton equation at feasible X:

$$X^{-1} \Delta X X^{-1} + \sum_{j=1}^{p} w_j A_i = X^{-1}, \qquad \mathbf{tr}(A_i \Delta X) = 0, \quad i = 1, \ldots, p$$

- follows from linear approximation $(X + \Delta X)^{-1} \approx X^{-1} - X^{-1} \Delta X X^{-1}$
- $n(n+1)/2 + p$ variables ΔX, w

16.3. Analytic Center of LMI

solution by block elimination

- eliminate ΔX from first equation: $\Delta X = X - \sum_{j=1}^{p} w_j X A_j X$
- substitute ΔX in second equation

$$\sum_{j=1}^{p} \mathbf{tr}(A_i X A_j X) w_j = b_i, \quad i = 1, \ldots, p$$

a dense positive definite set of linear equations with variable $w \in \mathbf{R}^p$

flop count (dominant terms) using Cholesky factorization $X = LL^T$:

- form p products $L^T A_j L$: $(3/2)pn^3$
- form $p(p+1)/2$ inner products $\mathbf{tr}((L^T A_i L)(L^T A_j L))$: $(1/2)p^2 n^2$
- solve (2) via Cholesky factorization: $(1/3)p^3$

16.4. 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/newton.pdf