文章编号:1000-582x(2001)05-0153-04

求解带二次简单约束的大型二次规划问题*

胡国雷，杨振华
（南京邮电学院 应用数理系，南京 210003）

摘　要:给出了一个求解形如$\frac{1}{2}x^{T}Hx+c^{T}x=\min$, $s.t.\ \|x\|_2\leqslant a$的二次规划问题的方法，该方法是由共轭斜量法(CG)和投影收缩算法(PC)的隐式方法组合而成的。对无约束问题，首先以$x^0=0$作为初始点，用(CG)方法进行求解，如果$\|x^k\|_2<a(k=1,2,\cdots)$，则原约束问题的解已经得到；否则用(CG)方法产生的迭代点的模一旦大于a，则以此点为新的初始点，改用隐式(PC)方法进行求解。数值例子的结果显示，该算法对处理大规模问题高效的，并且可大大提高精度。

关键词:二次规划；共轭斜量法；投影收缩方法
中图分类号:O 221.1　　**文献标识码**:A

考虑如下的带二次约束的二次规划问题

$$\frac{1}{2}\boldsymbol{x}^{T}\boldsymbol{H}\boldsymbol{x}+\boldsymbol{c}^{T}\boldsymbol{x}=\min$$
$$s.t.\ \|\boldsymbol{x}\|_2\leqslant a \tag{1}$$

其中$\boldsymbol{H}\in R^{n\times m}$，是一个半正定对称矩阵，$\boldsymbol{c}\in R^n$，$a$是一个确定的参数，有关求解问题(1)的一种基本方法是构造Lagrange函数：

$$L(x,\lambda)=\boldsymbol{x}^{T}\boldsymbol{H}\boldsymbol{x}+2\boldsymbol{c}^{T}\boldsymbol{x}+\lambda(\boldsymbol{x}^{T}\boldsymbol{x}-a^2) \tag{2}$$

由　$\nabla_x L(\boldsymbol{x},\lambda)=0,\ \nabla_\lambda L(\boldsymbol{x},\lambda)=0,$

得

$$\begin{cases}\boldsymbol{H}\boldsymbol{x}+\boldsymbol{c}+\lambda\boldsymbol{x}=0\\ \|\boldsymbol{x}\|=a\end{cases} \tag{3}$$

即$\begin{cases}(\boldsymbol{H}+\lambda\boldsymbol{I})\boldsymbol{x}+\boldsymbol{c}=0\\ \|\boldsymbol{x}\|=a\end{cases}$

从而有 $\|(\boldsymbol{H}+\lambda\boldsymbol{I})^{-1}\boldsymbol{c}\|=a$

令　$\varphi(\lambda)=\|(\boldsymbol{H}+\lambda\boldsymbol{I})^{-1}\boldsymbol{c}\|,\ S(\lambda)=(\boldsymbol{H}+\lambda\boldsymbol{I})^{-1}\boldsymbol{c}$

则　$\varphi^2(\lambda)=\boldsymbol{S}^{T}\boldsymbol{S}=\boldsymbol{c}^{T}(\boldsymbol{H}+\lambda\boldsymbol{I})^{-2}\boldsymbol{c}=a^2$

即
$$\varphi(\lambda)-a=0 \tag{4}$$

求出方程(4)的解λ^*，则得问题(1)的解

$$x^*=-(\boldsymbol{H}+\lambda^*\boldsymbol{I})^{-1}\boldsymbol{c}$$

牛顿法的求解过程一般需要求解8～9次线性方程组，所以它仅适用求解小规模问题，而对大型矩阵代价太大。因此人们还在继续寻找更有效的计算方法。

对大型的稀疏矩阵问题Golub和Von Matt[1]给出了一个方法，采用一系列不完全分解，产生一个含有上下界的Lagrange乘子序列，通过求解一个线性方程组，可求得界中的近似解$\boldsymbol{x}$。何[2]提出了一个求解问题(1)的有效方法，该方法是以$\boldsymbol{x}_0=0$作为初始点，如果$a\geqslant\|\boldsymbol{H}^{+}\boldsymbol{c}\|$，则仅采用(CG)方法求解该问题；如果$a\leqslant\|\boldsymbol{H}^{+}\boldsymbol{c}\|$，用此方法求解，一旦迭代点$\|\boldsymbol{x}^k\|>a$，则改用显式(PC)方法，本文提出的方法，前半部分与何[2]中(CG)方法相同，而后半部分将何[1]显式(PC)方法改为隐式方法，从后面数值例子的结果显示，隐式方法与Golub和Von Matt[1]方法、何[2]的显式方法相比计算效率更高，对坏条件问题的处理更有效，且可提高求解的精确度。

第1节中给出了问题(1)的等价投影方程；第2节介绍了(CG)方法和(PC)方法的隐式方法，并给出了算法的收敛性定理；第3节给出了算法的具体实现；最后绝对化出了一个大型二次规划的算例。

1　等价的投影方程

问题(1)的Lagrange函数为：

$$L(x,\lambda)=\boldsymbol{x}^{T}\boldsymbol{H}\boldsymbol{x}+2\boldsymbol{c}^{T}\boldsymbol{x}+\lambda(\boldsymbol{x}^{T}\boldsymbol{x}-a^2)$$

由凸规划的Kuhn-Tucker定理知，如果存在唯一的$\lambda^*\geqslant 0$及$\boldsymbol{x}^*$，使得$(\boldsymbol{x}^*,\lambda^*)$满足

$$\boldsymbol{H}\boldsymbol{x}+\boldsymbol{c}+\lambda\boldsymbol{x}=0$$

且
$$\lambda\geqslant 0,\boldsymbol{x}^{T}\boldsymbol{x}\leqslant a^2$$

* 收稿日期:2000-09-15
作者简介:胡国雷(1962-)，男，江苏武进人。南京邮电学院讲师，主要研究方向为优化与统计。

$$\lambda(x^{T}x - a^{2}) = 0 \tag{5}$$

则 x^* 是问题(1)的解。

设 $\Omega = \{x \in R^n, \|x\|_2 \leq a\}$

$P_\Omega(\cdot)$ 表示在凸集 Ω 上的投影算子,即

$$P_\Omega(x) = \begin{cases} x, & \|x\| \leq a \\ \dfrac{ax}{\|x\|}, & \|x\| > a \end{cases}$$

则(5)可写成如下的等价投影方程

$$x = P_\Omega(x + \lambda x) \tag{6}$$

将 $\lambda x = -(Hx + c)$ 代入投影方程,得到下列线性投影方程(称为 LPE),

$$\text{(LPE)} \qquad x = P_\Omega[x - (Hx + c)] \tag{7}$$

因此,求解问题(1)等价于求解如下残差函数的零点:

$$e(x) := x - P_\Omega[x - (Hx + c)] \tag{8}$$

引入调比因子 μ 得

$$e(x,\mu) := x - P_\Omega[x - \mu(Hx + c)] \tag{9}$$

2 共轭斜量法和投影收缩的隐式方法

1) 共轭斜量法

共轭斜量法求解凸的二次函数的极小值是一个精确的线性搜寻算法

$$x^{k+1} = x^k - \alpha_k s^k$$

$$s^0 = g^0, \beta_0 = 0$$

$$s^{k+1} = g^{k+1} + \beta_k s^k$$

$$\beta_k = \frac{g^{k+1^T} g^{k+1}}{g^{k^T} g^k}$$

$$\alpha_k = \frac{g^{kT} g^k}{s^{k^T} H s^k}$$

且 $\qquad g = Hx + c$

如果 H 是对称半正定矩阵,(CG)方法求解 $Hx + c = 0$ 经 $m(m \leq n)$ 步后停止。且对所有 $i(1 \leq i \leq m)$,有下列关系式成立

$$s^{i^T} H s^j = 0$$

$$g^{i^T} g^j = 0 \quad j = 0,1,\cdots,i-1$$

$$g^{i^T} s^j = 0$$

定理 1 若将 $\tilde{x}$ 表示 $H^+ c$,则对 $k(k \leq m)$ 有

$$\|x^{k+1} - \tilde{x}\| < \|\tilde{x}^k - \tilde{x}\|$$

证明见[3]。

何[2]给出了有关(CG)方法的另一个性质

定理 2 若(CG)方法的初始点为 $x^0 = 0$,则对所有 $k(k < m)$,有

$$\|x^{k+1}\| < \|x^k\|$$

2) 投影收缩算法的隐式方法

设 $\gamma \in (0,2)$ 是一个常数,G 是一个正定矩阵,给定 $x^0 \in R^n$,对 $k = 0,1,\cdots$,如果 $x^0 \notin \Omega^n$,则求解方程 $Q_k(x) = 0$ 得 x^{k+1},其中

$$Q_k(x) = x + \mu(Hx + c) - x^k - \mu(Hx^k + c) + \gamma\rho_k G^{-1} e(x^k,\mu) \tag{10}$$

$$\rho_k = \frac{\|e(x^k,\mu)\|^2}{e(x^k,\mu)^T G^{-1} e(x^k,\mu)} \tag{11}$$

若 $G = I$,则

$$x^{k+1} = x^k + \mu(Hx^k + c) - \mu(Hx^{k+1} + c) - \gamma e(x^k,\mu)$$

即 $$x^{k+1} = x^k - \gamma(I + \mu H)^{-1} e(x^k,\mu) \tag{12}$$

3) 收敛性

定理 3 设 $x^* \in \Omega^*$,则

$$\{(I + \mu H)(x - x^*)\}^T e(x^k,\mu) \geq \|e(x^k,\mu)\|^2 + \mu^2 \|x - x^*\|_H^2 \tag{13}$$

由 FI1 + FI2(何[4])即可证得。

定理 4

$$\|(I + \mu H)(x^{k+1} - x^*\|_G^2 \leq \|(I + \mu H)(x^k - x^*)\|_G^2 - \gamma(2-\gamma)\rho_k \|e(x^k,\mu)\|^2 - 2\gamma\rho_k\mu^2 \|x^k - x^*\|_H^2 \tag{14}$$

证:由

$$x^{k+1} - x^k + \mu H x^{k+1} - \mu H x^k = -\gamma\rho_k G^{-1} e(x^k,\mu)$$

$$(I + \mu H)(x^{k+1} - x^k) = -\gamma\rho_k G^{-1} e(x^k,\mu)$$

$$\begin{aligned}
&\|(I + \mu H)(x^{k+1} - x^k)\|_G^2 = \\
&\|(I + \mu H)[x^{k+1} - x^k + x^k - x^*]\|_G^2 = \\
&\|(I + \mu H)(x^k - x^*) - \gamma\rho_k G^{-1} e(x^k,\mu)\|_G^2 = \\
&\|(I + \mu H)(x^k - x^*)\|_G^2 - \\
&2\gamma\rho_k \{(I + \mu H)(x^k - x^*)\}^T e(x^k,\mu) + \\
&\gamma^2\rho_k^2 e(x^k,\mu) G^{-1} e(x^k,\mu) \leq \\
&\|(I + \mu H)(x^k - x^*)\|_G^2 - \\
&2\gamma\rho_k \{\|e(x^k,\mu)\|^2 + \mu^2 \|x^k - x^*\|_H^2\} + \\
&\gamma^2\rho_k \|e(x^k,\mu)\|^2 = \quad \text{(利用(13)和(11)式)} \\
&\|(I + \mu H)(x^k - x^*\|_G^2 - \\
&\gamma(2-\gamma)\rho_k \|e(x^k,\mu)\|^2 - 2\gamma\rho_k\mu^2 \|x^k - x^*\|_H^2
\end{aligned}$$

定理 5 由投影收缩算法的隐式方法产生的序列 $\{x^k\}$ 收敛于解点 x^*。

证:由定理 2 得

$$\sum_{k=0}^{\infty} \gamma(2-\gamma)\rho_k \|e(x^k,\mu)\|^2 \leq \|(I + \mu H)(x^0 - \tilde{x})\|_G^2$$

即有 $$\lim_{k\to\infty} \|e(x^k,\mu)\|^2 = 0$$

且$\{x^k\}$有界，所以$\{x^{k_j}\}$为收敛于x^*的子序列，又因为

$\|(I+\mu H)(x^{k+1}-x^*\|_G^2 \le \|(I+\mu H)(x^k-x^*\|_G^2$

所以有

$$\lim_{k\to\infty} x^k = x^*$$

3 关于算法的具体实现

步骤 1：

$x^0=0$，当$\|x^l\|\le a$且$l<n$，由(CG)方法计算出x^{l+1}，如果$\|x^l\|\le a$，$l=0,1,\cdots,n$，用(CG)方法求解线性方程$Hx+c=0$产生的迭代点x^n是问题(1)的解；否则一旦$\|x^l\|>a$，$(l\le n)$，取$\tilde{x}^0=\frac{ax^l}{\|x\|^l}$作为新的初始点，并且转用投影收缩算法的隐式方法(12)，在(12)中，取$\mu=\frac{a}{\|H\tilde{x}^0+c\|}$；

步骤 2：

$$x^0=\tilde{x}^0$$

$$e(x,\mu):=x-P_\Omega[x-\mu(Hx+c)]$$

$$x^{k+1}=x^k-\gamma(I+\mu H)^{-1}e(x^k,\mu) \quad k=0,1,\cdots$$

取停机准则为

$$\max\left\{\left|\frac{\|x\|-a}{a}\right|,\frac{\|e(x)\|}{\sqrt{a\|c\|}}\right\}\le\varepsilon$$

归结起来，算法如下

```
给定 K 和 ε
done := false
  l := 0
  x := 0
while not ‖x‖ < a  and  l < n  do
(用(CG)步计算一个新的迭代变量 x)
  l := l + 1
end
ifl := n   then done := true
else
  k := 0
  x := ax/‖x‖
  μ := a/‖Hx + c‖
end
while not done and k ≤ K do
  e(x,μ) := x - P_Ω[x - μ(Hx + c)]
  x := x - 1.8(I + μH)^{-1} e(x,μ)
  k := k + 1
if  max{|(‖x‖ - a)/a|, ‖e(x)‖/√(a‖c‖)} ≤ ε
  done := true
end
```

4 数值实验结果

这一节我们给出一个问题(1)的大规模的数值例子

设$H=A^TA$，$c=-A^Tb$，$A:=U\sum V^T$，其中

$U=I_m-2\frac{uu^T}{\|u\|_2^2}$，

$V=I_n-2\frac{vv^T}{\|v\|_2^2}$是豪斯荷尔德矩阵，

$\sum=\mathrm{diag}(\sigma_k)$是一个$m\times n$对角阵，

向量u，v和b是同余伪随机数：

$u_1=13\,846+u_i=(3\,146u_{i-1}+13\,846)\bmod 46\,216$ $i=2,\cdots,m$

$v_1=13\,846$

$v_j=(42\,108v_{j-1}+13\,846)\bmod 46\,273$ $j=2,\cdots,n$

$b_1=13\,846$

$b_i=(45\,278b_{i-1}+13\,846)\bmod 46\,219$ $i=2,\cdots,m$

取$\varepsilon_1=5\times10^{-6}$(与[6]相同)，$\varepsilon_2=5\times10^{-12}$。

文中算法用 Matlab 语言实现，表中 l 表示用(CG)方法直到$\|x^l\|>a$的迭代步数，k表示用(PC)方法直到满足停机准则的迭代次数。

例：取$m=2\,000$，$n=1\,000$，$\sigma_k=\cos\frac{k\pi}{n+1}+1$，$k=1,\cdots,n$，

$\mathrm{cond}(H)=\mathrm{cond}(A^TA)=1.65\times10^{11}$，$\|H^+c\|=4.37\times10^9$

表 1 给出了$a=10^4,\cdots,10^7$的结果。

表 1 $a=10^4,\cdots,10^7$ 的结果

m	n	$\|H^+c\|$	a	l	μ	k ($\varepsilon_1=5\times10^{-6}$)	k ($\varepsilon_2=5\times10^{-12}$)
2 000	1 000	4.37×10^9	10^4	1	0.01	22	77
2 000	1 000	4.37×10^9	10^5	1	0.14	12	23
2 000	1 000	4.37×10^9	10^6	7	11.05	11	19

m	n	$\|H^{+}c\|$	a	1	μ	k ($\varepsilon_1 = 5\times10^{-6}$)	k ($\varepsilon_2 = 5\times10^{-12}$)
2 000	1 000	4.37×10^{9}	2×10^{6}	20	47.17	13	29
2 000	1 000	4.37×10^{9}	3×10^{6}	36	97.13	18	39
2 000	1 000	4.37×10^{9}	5×10^{6}	63	215.29	24	58
2 000	1 000	4.37×10^{9}	8×10^{6}	111	573.93	24	69
2 000	1 000	4.37×10^{9}	10^{7}	146	867.98	31	79

注:在(PC)方法中适当选择参数 μ 对收敛速度能起到很关键的作用,对同样的问题,如果选择 $\mu = 1$,则迭代的次数比原来要多出许多倍,特别在 a 取值较大时迭代次数差别更大,见表2

表2 迭代次数差别

m	n	$\|H^{+}c\|$	a	1	μ	k ($\varepsilon_1 = 5\times10^{-6}$)	k ($\varepsilon_2 = 5\times10^{-12}$)
2 000	1 000	4.37×10^{9}	10^{2}	1	1	34	137
2 000	1 000	4.37×10^{9}	10^{5}	1	1	19	53
2 000	1 000	4.37×10^{9}	10^{6}	7	1	73	237
2 000	1 000	4.37×10^{9}	2×10^{6}	20	1	410	1 629
2 000	1 000	4.37×10^{9}	3×10^{6}	36	1	992	4 513
2 000	1 000	4.37×10^{9}	5×10^{6}	63	1	2 761	15 009
2 000	1 000	4.37×10^{9}	8×10^{6}	111	1	2 718	> 20 000
2 000	1 000	4.37×10^{9}	10^{7}	146	1	10 210	> 20 000

参考文献:

[1] G H Golub U VON MATT. Quadratically constrained least squares and quadratic problems[J]. Numerische Mathematical, 1990(59):561 - 580.

[2] B S HE. Solving trust region problem in large scale optimization [J]. Journal of Computational Mathematics, 2000, 18(1):1 - 12.

[3] M HESTENES. Conjugate Direction Methods in optimization [M]. New York: Springer - Verlag, Berlin; Heidelberg, 1980.

[4] B S HE. A new method for a class of linear variational inequalities[J] Mathematical Programming, 1994(66):137 - 144.

[5] B S HE. A projection and contraction method for a class of linear complementary problems and its application in convex quadratic programming [J]. Applied Mathematics and Optimization, 1992, 25:247 - 262.

[6] B S HE. Solving a class of linear projection equations [J]. Numerische Mathematical, 1994, 68:71 - 80.

Solving Large-scale Quadratic Programming with Simple Quadratic Constraint

HU Guo-lei, YANG Zhen-hua

(Department of Applied Mathematics and Physics, Nanjing University of Posts & Telecommunications, Nanjing 210003, China)

Abstract: A method soving the quadratic programming $\frac{1}{2}x^{T}Hx + c^{T}x = \min,\ s.t.\ \|x\|_2 \leqslant a$ is established, which is combined by conjugate gradient method and implicit projection and contraction method. The unconstrained problem is solved by conjugate gradient method with the initial point $x^0 = 0$. The solution of the constrained problem is obtained if $\|x^k\|_2 < a$ ($k = 1, 2, \cdots$), otherwise, once the norm of the interior point is greater than a, implicit PC method is used, started with this point. The numerical results show that this algorithm is very effective for large-scale problem, and the precision is improved.

Key words: quadratic programming; conjugate gradient method; projection and contraction method

(责任编辑 张小强)