第38卷第9期 同济大学学报(自然科学版) Vol.38 No.9
2010年9月 JOURNAL OF TONGJI UNIVERSITY(NATURAL SCIENCE) Sep. 2010

文章编号：0253-374X(2010)09-1387-05 DOI:10.3969/j.issn.0253-374x.2010.09.026

新的拉格朗日乘子方法

濮定国[1]，金　中[1,2]

(1.同济大学 数学系，上海 200092；2.上海海事大学 数学系，上海 200135)

摘要：对于约束优化问题，提出一类新的结合 Fischer-Burmeister 非线性互补(NCP)函数的增广拉格朗日函数，它的无约束极小解对应于原约束问题(NLP)的解及其乘子；同时提出相对应的拉格朗日乘子方法.该方法可实现并具有全局收敛性.

关键词：约束优化；非线性互补函数；拉格朗日函数；乘子；收敛性

中图分类号：O 221.2 **文献标识码**：A

New Lagrangian Multiplier Methods

PU Dingguo[1], *JIN Zhong*[1,2]

(1. Department of Mathematics, Tongji University, Shanghai 200092, China; 2. Department of Mathematics, Shanghai Maritime University, Shanghai 200135, China)

Abstract: A new class of augmented Lagrangian functions with the Fischer-Burmeister NCP function and a Lagrangian multiplier method are proposed for the minimization of a smooth function subject to smooth equation and inequality constraints. This method is based on the solutions of the unconstrained optimization which is a reformulation of the primal constrained problem. These methods are implementable and globally convergent.

Key words: constrained optimization; nonlinear complementarity function; Lagrangian function; multiplier; convergence

考虑非线性约束优化问题(NLP)：min $f(\boldsymbol{x})$，s.t. $H(\boldsymbol{x})=0$，$G(\boldsymbol{x})\leqslant 0$，其中 $f:\mathbf{R}^n\to\mathbf{R}$，$H(\boldsymbol{x})=(h_1(\boldsymbol{x}),h_2(\boldsymbol{x}),\cdots,h_p(\boldsymbol{x}))^{\mathrm{T}}:\mathbf{R}^n\to\mathbf{R}^p$ 和 $G(\boldsymbol{x})=(g_1(\boldsymbol{x}),g_2(\boldsymbol{x}),\cdots,g_m(\boldsymbol{x}))^{\mathrm{T}}:\mathbf{R}^n\to\mathbf{R}^m$ 是二次连续可微函数.

用 X 来表示问题(NLP)的可行集.问题(NLP)的拉格朗日函数为

$$L(\boldsymbol{x},\boldsymbol{\omega},\boldsymbol{\lambda})=f(\boldsymbol{x})+\boldsymbol{\omega}^{\mathrm{T}}H(\boldsymbol{x})+\boldsymbol{\lambda}^{T}G(\boldsymbol{x})\qquad(1)$$

其中 $\boldsymbol{\omega}=(\omega_1,\omega_2,\cdots,\omega_p)^{\mathrm{T}}\in\mathbf{R}^p$ 和 $\boldsymbol{\lambda}=(\lambda_1,\lambda_2,\cdots,\lambda_m)^{\mathrm{T}}\in\mathbf{R}^m$ 是乘子向量.为方便起见，用$(\boldsymbol{x},\boldsymbol{\omega},\boldsymbol{\lambda})$表示列向量$(\boldsymbol{x}^{\mathrm{T}},\boldsymbol{\omega}^{\mathrm{T}},\boldsymbol{\lambda}^{\mathrm{T}})^{\mathrm{T}}$.

Karush-Kuhn-Tucker(KKT)点$(\bar{\boldsymbol{x}},\bar{\boldsymbol{\omega}},\bar{\boldsymbol{\lambda}})\in\mathbf{R}^n\times\mathbf{R}^p\times\mathbf{R}^m$ 是指满足问题(NLP)的一阶最优必要条件(2)的点，即

$$\nabla_x L(\bar{\boldsymbol{x}},\bar{\boldsymbol{\omega}},\bar{\boldsymbol{\lambda}})=0,\ G(\bar{\boldsymbol{x}})\leqslant 0,\ H(\bar{\boldsymbol{x}})=0,$$
$$\bar{\boldsymbol{\lambda}}_i^{T}g_i(\bar{\boldsymbol{x}})=0,\ 0\leqslant i\leqslant m,\ \bar{\boldsymbol{\lambda}}\geqslant 0.\qquad(2)$$

解非线性约束优化问题的一类重要方法是用一系列无约束子问题替代原约束问题求解.这种方法叫做乘子方法或增广拉格朗日方法[1-6]，它的无约束问题里的目标函数 $S(\boldsymbol{x},\boldsymbol{\omega},\boldsymbol{\lambda},\boldsymbol{C})$称做增广拉格朗日函数，其中 $\boldsymbol{C}$ 是正参数或者正参向量.经典的增广拉格朗日方法中使用的极大函数可能在无数个点处不可微.为了克服这一缺点，本文结合 Fischer-Burmeister NCP 函数[7-9]提出了新的增广拉格朗日函数以及一种拉格朗日乘子方法.这一方法基于求解与原约束问题等价的无约束优化问题.该算法可实现并具有全局收敛性.

Fischer-Burmeister 函数 ψ 具有非常简单的结构

$$\psi(a,b)=\sqrt{a^2+b^2}-a-b$$

ψ 除了在原点外处处连续可微，且它在原点处是强半光滑的.令

$$\psi_i(\boldsymbol{x},\lambda_i,C_i)=\psi((-C_ig_i(\boldsymbol{x})),\lambda_i),\quad 1\leqslant i\leqslant m$$

其中 $C_i>0$，$\Phi(\boldsymbol{x},\boldsymbol{\lambda},\boldsymbol{C})=(\phi_1(\boldsymbol{x},\lambda_1,C_1),\cdots,\phi_m(\boldsymbol{x},\lambda_m,C_m))^{\mathrm{T}}$.显然，KKT 点条件(2)等价于条件 $\Phi(\boldsymbol{x},\boldsymbol{\lambda},\boldsymbol{C})=0$，$H(\boldsymbol{x})=0$ 和$\nabla_x L(\boldsymbol{x},\boldsymbol{\omega},\boldsymbol{\lambda})=0$.

收稿日期：2009-05-22
基金项目：国家自然科学基金资助项目(10771162)
作者介绍：濮定国(1948—)，男，教授，博士生导师，主要研究方向为最优化理论与方法.E-mail:madpu@tongji.edu.cn

1 预备知识

假定 1 函数 $f, h_i(\boldsymbol{x}), i=1,2,\cdots,p$, 和 $g_j(\boldsymbol{x}), j=1,2,\cdots,m$, 是二次 Lipschitz 连续可微的.

对于问题(NLP),结合 Fischer-Burmeister NCP 函数定义拉格朗日乘子函数 $S:\mathbf{R}^{n+m+p}\to\mathbf{R}$ 如下:

$$\min S(\boldsymbol{x},\boldsymbol{\omega},\boldsymbol{\lambda},\boldsymbol{C},\boldsymbol{D}) \overset{\text{def}}{=} f(\boldsymbol{x}) + \boldsymbol{\omega}^{\mathrm{T}}H(\boldsymbol{x}) + \sum_{i=1}^{p} D_i(h_i(\boldsymbol{x}))^2/2 + \sum_{i=1}^{m}[(\phi_i(\boldsymbol{x},\lambda_i,C_i)+\lambda_i)^2-(\lambda_i)^2]/(2C_i) \tag{3}$$

其中 $\boldsymbol{\omega}=(\omega_1,\omega_2,\cdots,\omega_p)^{\mathrm{T}}$ 和 $\boldsymbol{\lambda}=(\lambda_1,\lambda_2,\cdots,\lambda_m)^{\mathrm{T}}$ 是拉格朗日乘子, $\boldsymbol{C}=(C_1,C_2,\cdots,C_m)>0$ 和 $\boldsymbol{D}=(D_1,D_2,\cdots,D_p)>0$ 是参数.

定义指标集 I_0 和 I_1 为: $I_1(\boldsymbol{x},\boldsymbol{\lambda})=\{i\,|\,(g_i(\boldsymbol{x}),\lambda_i)\neq(0,0), i=1,2,\cdots,m\}$; $I_0(\boldsymbol{x},\boldsymbol{\lambda})=\{i\,|\,(g_i(\boldsymbol{x}),\lambda_i)=(0,0), i=1,2,\cdots,m\}$. 由 $\phi_j(\boldsymbol{x},\lambda_j,C_j)$ 的定义,可得对 $j\in I_0(\boldsymbol{x},\boldsymbol{\lambda})$, $\nabla[(\phi_j(\boldsymbol{x},\lambda_j,C_j)+\lambda_j)^2]=0$, 对 $j\in I_1(\boldsymbol{x},\boldsymbol{\lambda})$

$$\nabla[(\phi_j(\boldsymbol{x},\lambda_j,C_j)+\lambda_j)^2]/(2C_j) = \left(\frac{C_jg_j(\boldsymbol{x})}{\sqrt{(C_jg_j(\boldsymbol{x}))^2+\lambda_j^2}}+1\right)\left(\sqrt{(C_jg_j(\boldsymbol{x}))^2+\lambda_j^2}+C_jg_j(\boldsymbol{x})\right)\nabla g_j(\boldsymbol{x}) \tag{4}$$

由 S 的定义和式(4),知

$$\nabla S(\boldsymbol{x},\boldsymbol{\omega},\boldsymbol{\lambda},\boldsymbol{C},\boldsymbol{D}) = \nabla L(\boldsymbol{x},\boldsymbol{\omega},\boldsymbol{\lambda}) + \sum_{j=1}^{p}D_j(h_j(\boldsymbol{x})\nabla h_j(\boldsymbol{x})) + \sum_{j=1}^{m}\phi_j(\boldsymbol{x},\lambda_j,C_j)\nabla g_j(\boldsymbol{x}) + \sum_{j\in I_1}\left(\frac{C_jg_j(\boldsymbol{x})}{\sqrt{(C_jg_j(\boldsymbol{x}))^2+\lambda_j^2}}\right)\left(\sqrt{(C_jg_j(\boldsymbol{x}))^2+\lambda_j^2}+C_jg_j(\boldsymbol{x})\right)\nabla g_j(\boldsymbol{x}) \tag{5}$$

$S(\boldsymbol{x},\boldsymbol{\omega},\boldsymbol{\lambda},\boldsymbol{C},\boldsymbol{D})$ 在问题(NLP)的 KKT 点 $(\bar{\boldsymbol{x}},\bar{\boldsymbol{\omega}},\bar{\boldsymbol{\lambda}})$ 处的海塞矩阵为

$$\nabla^2 S(\bar{\boldsymbol{x}},\bar{\boldsymbol{\omega}},\bar{\boldsymbol{\lambda}},\boldsymbol{C},\boldsymbol{D}) = \nabla^2 L(\bar{\boldsymbol{x}},\bar{\boldsymbol{\omega}},\bar{\boldsymbol{\lambda}}) + \sum_{j=1}^{p}D_j\nabla h_j(\boldsymbol{x})(\nabla h_j(\boldsymbol{x}))^{\mathrm{T}} + \sum_{j\in I_1}\left(2C_j + \frac{3C_j^2g_j(\bar{\boldsymbol{x}})}{\sqrt{(C_jg_j(\bar{\boldsymbol{x}}))^2+\bar{\lambda}_j^2}} - \frac{C_j^4(g_j(\bar{\boldsymbol{x}}))^3}{((C_jg_j(\bar{\boldsymbol{x}}))^2+\bar{\lambda}_j^2)^{3/2}}\right)\cdot \nabla g_j(\bar{\boldsymbol{x}})(\nabla g_j(\bar{\boldsymbol{x}}))^{\mathrm{T}} \tag{6}$$

定义 1 若点 $(\boldsymbol{x},\boldsymbol{\omega},\boldsymbol{\lambda})$ 满足一阶 KKT 条件,且对于所有的 $\boldsymbol{d}\in P(\boldsymbol{x})=\{\boldsymbol{d}\,|\,\boldsymbol{d}^{\mathrm{T}}\nabla h_i(\boldsymbol{x})=0, i=1,\cdots,p, \boldsymbol{d}^{\mathrm{T}}\nabla g_j(\boldsymbol{x})=0, j\in J(\boldsymbol{x})\}$, $J(\boldsymbol{x})=\{j\,|\,j=1,\cdots,m,\lambda_j>0\}$ 和 $\boldsymbol{d}\neq 0$ 有 $\boldsymbol{d}^{\mathrm{T}}\nabla_{xx}^2L(\boldsymbol{x},\boldsymbol{\omega},\boldsymbol{\lambda})\boldsymbol{d}>0$,则称该点满足问题(NLP)的强二次充分条件.

假定 2 强二次充分条件在问题(NLP)的任一 KKT 点都成立.

引理 1 若半正定 $n\times n$ 矩阵 $\boldsymbol{A}$ 和 $n\times n$ 矩阵 $\boldsymbol{B}$ 对任一 $\boldsymbol{d}\in\mathbf{R}^n$, $\boldsymbol{A}\,\boldsymbol{d}=0$, 满足 $\boldsymbol{d}^T\boldsymbol{B}\,\boldsymbol{d}>0$,则存在一个 m_1,使得对于任意的 $m>m_1$, $\boldsymbol{B}+m\boldsymbol{A}$ 是正定的.

引理 2 假设假定 1,2 成立.如果 $(\bar{\boldsymbol{x}},\bar{\boldsymbol{\omega}},\bar{\boldsymbol{\lambda}})$ 是问题(NLP)的一个 KKT 点,则对充分大的 $\boldsymbol{C},\boldsymbol{D}$, $S(\boldsymbol{x},\boldsymbol{\omega},\boldsymbol{\lambda},\boldsymbol{C},\boldsymbol{D})$ 在点 $(\bar{\boldsymbol{x}},\bar{\boldsymbol{\omega}},\bar{\boldsymbol{\lambda}})$ 是强凸函数.

引理 2 推出存在 $\varepsilon>0$ 和 $\bar{m}$,使得对所有的 $C_j\geqslant\bar{m}$, $1\leqslant j\leqslant m$, $D_j\geqslant\bar{m}$, $1\leqslant j\leqslant p$ 和 $\|\boldsymbol{x}\|=1$,且

$$\boldsymbol{x}^{\mathrm{T}}\nabla^2 S(\bar{\boldsymbol{x}},\bar{\boldsymbol{\omega}},\bar{\boldsymbol{\lambda}},\boldsymbol{C},\boldsymbol{D})\boldsymbol{x}\geqslant 2\varepsilon$$

同时由连续性可知,存在 $\eta>0$,使得对所有的 $(\boldsymbol{x}_1,\boldsymbol{\omega}_1,\boldsymbol{\lambda}_1)\in B_\eta(\bar{\boldsymbol{x}},\bar{\boldsymbol{\omega}},\bar{\boldsymbol{\lambda}},\boldsymbol{C},\boldsymbol{D})=\{\|(\boldsymbol{x}_1,\boldsymbol{\omega}_1,\boldsymbol{\lambda}_1)-(\bar{\boldsymbol{x}},\bar{\boldsymbol{\omega}},\bar{\boldsymbol{\lambda}})\|\leqslant\eta\}$,有

$$\boldsymbol{x}^{\mathrm{T}}\nabla^2 S(\boldsymbol{x}_1,\boldsymbol{\omega}_1,\boldsymbol{\lambda}_1,\boldsymbol{C},\boldsymbol{D})\boldsymbol{x}\geqslant\varepsilon>0 \tag{7}$$

2 算法

步 0:选择参数 $\boldsymbol{C}^0>0$, $\boldsymbol{D}^0>0$, $0\leqslant\eta\ll 1$, $0<\theta_1<1<\theta_2$.给出初始点 $\boldsymbol{x}_0\in\mathbf{R}^n$, $\boldsymbol{\omega}^0=(\omega_1^0,\omega_2^0,\cdots,\omega_p^0)\in\mathbf{R}^p$, $\boldsymbol{\lambda}^0=(\lambda_1^0,\lambda_2^0,\cdots,\lambda_m^0)\in\mathbf{R}^m$.令 $k=0$.

步 1:解式(8)得 $\boldsymbol{x}^{k+1}$(见 4.1,4.2),

$$\min S(\boldsymbol{x},\boldsymbol{\omega}^k,\boldsymbol{\lambda}^k,\boldsymbol{C}^k,\boldsymbol{D}^k) \overset{\text{def}}{=} f(\boldsymbol{x}) + (\boldsymbol{\omega}^k)^{\mathrm{T}}H(\boldsymbol{x}) + \sum_{i=1}^{p}D_i^{\,k}(h_i(\boldsymbol{x}))^2/2 + \sum_{i=1}^{m}[(\phi_i(\boldsymbol{x},\lambda_i^k,C_i^{\,k})+\lambda_i^k)^2-(\lambda_i^k)^2]/(2C_i^k) \tag{8}$$

如果 $\|H(\boldsymbol{x}^{k+1})\|_\infty\leqslant\eta$ 且 $\|\Phi(\boldsymbol{x}^{k+1},\boldsymbol{\lambda}^k,\boldsymbol{C}^k)\|_\infty\leqslant\eta$,停止.

步 2:对于 $i=1,2,\cdots,p$,如果 $|h_i(\boldsymbol{x}^{k+1})|\leqslant\theta_1|h_i(\boldsymbol{x}^k)|$,则 $D_i^{k+1}=D_i^k$,否则 $D_i^{k+1}=\theta_2D_i^k$.

对于 $i=1,2,\cdots,m$,如果 $|\phi_i(\boldsymbol{x}^{k+1},\lambda_i^k,C_i^k)|\leqslant\theta_1|\phi_i(\boldsymbol{x}^k,\lambda_i^k,C_i^k)|$,则 $C_i^{k+1}=C_i^k$,否则 $C_i^{k+1}=\theta_2C_i^k$.

步 3:计算 $\boldsymbol{\omega}^{k+1}$ 和 $\boldsymbol{\lambda}^{k+1}$ 得

$$\omega_i^{k+1}=\omega_i^k+D_i^{\,k}h_i(\boldsymbol{x}^{k+1}) \tag{9}$$

$$\lambda_i^{k+1}=\left(\sqrt{(C_i^kg_i(\boldsymbol{x}^{k+1}))^2+(\lambda_i^k)^2}+C_i^kg_i(\boldsymbol{x}^{k+1})\right)\geqslant 0 \tag{10}$$

步 4:令 $k=k+1$,转步 1.

3 收敛性

在下面讨论中取 $\eta=0$.令 $\mu_i^k=\omega_i^{\,k}/\sqrt{D_i^k}$ 和

$\nu_i^k=\lambda_i^k/\sqrt{C_i^k}$，$\boldsymbol{\mu}^k=(\mu_1^k,\mu_2^k,\cdots,\mu_p^k)^{\mathrm{T}}$ 和 $\boldsymbol{v}^k=(\nu_1^k,\nu_2^k,\cdots,\nu_m^k)^{\mathrm{T}}$.

引理 3　假设可行集 X 非空. 如果 $f(\boldsymbol{x})$ 有下界，则存在 τ 使得 $\|\boldsymbol{\mu}^k\|^2+\|\boldsymbol{v}^k\|^2\leqslant\tau k$.

证明　对于任意 k

$$S(\boldsymbol{x}^{k+1},\boldsymbol{\omega}^k,\boldsymbol{\lambda}^k,\boldsymbol{C}^k,\boldsymbol{D}^k)-f(\boldsymbol{x}^{k+1})=\sum_{i=1}^{p}[(\omega_i^{k+1})^2-(\omega_i^k)^2]/(2D_i^k)+\sum_{i=1}^{m}[(\lambda_i^{k+1})^2-(\lambda_i^k)^2]/(2C_i^k)\tag{11}$$

另一方面，对任意 $\bar{\boldsymbol{x}}\in X$（可行集），可得 $H(\bar{\boldsymbol{x}})=0$，$g(\bar{\boldsymbol{x}})\leqslant 0$，$0\leqslant\phi_i(\bar{\boldsymbol{x}},\lambda_i^k,C_i^k)+\lambda_i^k\leqslant\lambda_i^k$.

$$S(\bar{\boldsymbol{x}},\boldsymbol{\omega}^k,\boldsymbol{\lambda}^k,\boldsymbol{C}^k,\boldsymbol{D}^k)-f(\bar{\boldsymbol{x}})=\sum_{i=1}^{m}[(\phi_i(\bar{\boldsymbol{x}},\lambda_i^k,C_i^k)+\lambda_i^k)^2-(\lambda_i^k)^2]/(2C_i^k)\leqslant 0\tag{12}$$

所以

$$\|\boldsymbol{\mu}^{k+1}\|^2+\|\boldsymbol{v}^{k+1}\|^2\leqslant\|\boldsymbol{\mu}^k\|^2+\|\boldsymbol{v}^k\|^2+2[f(\bar{\boldsymbol{x}})-f(\boldsymbol{x}^{k+1})]\tag{13}$$

定理 1　假设可行集 X 非空. 执行算法，要么它在有限步迭代后终止，要么

$$\liminf_{k\to\infty}f(\boldsymbol{x}^k)\to-\infty$$

证明　假设此定理非真，则算法不会有限步迭代终止且 $\liminf\limits_{k\to\infty}f(\boldsymbol{x}^k)>-\infty$. 令

$$J_e=(i\mid\lim_{k\to\infty}h_i(\boldsymbol{x}^k)\neq 0)\quad \bar{J}_e=(i\mid\lim_{k\to\infty}D_i^k=\infty)$$

$$J_i=(i\mid\lim_{k\to\infty}\phi_i(\boldsymbol{x}^k)\neq 0)\quad \bar{J}_i=(i\mid\lim_{k\to\infty}\mid C_i^k=\infty)$$

由假设可知：$J_e\cup J_i$ 或 $\bar{J}_e\cap\bar{J}_i$ 非空.

对任意 $\bar{\boldsymbol{x}}\in X$，有

$$f(\bar{\boldsymbol{x}})+\sum_{i=1}^{m}[(\phi_i(\bar{\boldsymbol{x}},\lambda_i^k,C_i^k)+\lambda_i^k)^2-(\lambda_i^k)^2]/(2C_i^k)\geqslant f(\boldsymbol{x}^{k+1})+(\boldsymbol{\omega}^k)^{\mathrm{T}}H(\boldsymbol{x}^{k+1})+\sum_{i=1}^{p}[D_ih_i(\boldsymbol{x}^{k+1})^2/2+\sum_{i=1}^{m}[(\phi_i(\boldsymbol{x}^{k+1},\lambda_i^k,C_i^k)+\lambda_i^k)^2-(\lambda_i^k)^2]/(2C_i^k)\tag{14}$$

所以 $C_i^{k+1}\geqslant C_i^k$ 和 $D_i^{k+1}\geqslant D_i^k$ 说明

$$f(\bar{\boldsymbol{x}})-f(\boldsymbol{x}^{k+1})\geqslant O(k)+\sum_{i\in\bar{J}_e}D_i^k[((h_i(\boldsymbol{x}^{k+1})^2+\omega_i^k/D_i^k)^2-(\omega_i^k/D_i^k)^2/2+\sum_{i\in\bar{J}_i}C_i^k[(\phi_i(\boldsymbol{x}^{k+1},\lambda_i^k,C_i^k)/C_i^k+\lambda_i^k/C_i^k)^2-(\lambda_i^k/C_i^k)^2]/2\tag{15}$$

因为算法不会有限步终止，所以对任一 $\bar{k}$ 存在一个 $k>\bar{k}$，使得要么对于某一 $i\in\bar{J}_e$ 和 $\tau_1>0$，$D_i^{k+1}>D_i^k$ 且 $h_i(\boldsymbol{x}^{k+1})>\tau_1$，要么对于某一 $i\in\bar{J}_i$ 和 $\tau_1>0$，$C_i^{k+1}>C_i^k$ 且 $g_i(\boldsymbol{x}^{k+1})>\tau_1$，其中 $\tau_1>0$. 可知对充分大的 k 要么对某一 $i\in\bar{J}_e$，有

$$D_i^k[((h_i(\boldsymbol{x}^{k+1})^2+\omega_i^k/D_i^k)^2-(\omega_i^k/D_i^k)^2/2>D_i^k\tau_1{}^2/2>k^2\tau_1^2/2$$

要么对某一 $i\in\bar{J}_i$，有

$$C_i^k[(\phi_i(\boldsymbol{x}^{k+1},\lambda_i^k,C_i^k)/C_i^k+\lambda_i^k/C_i^k)^2-(\lambda_i^k/C_i^k)^2]/2>C_i^k\tau_1^2/2>k^2\tau_1^2/2$$

则下面的式子成立：

$$f(\bar{\boldsymbol{x}})-f(\boldsymbol{x}^{k+1})\geqslant k^2\tau_1^2/4$$

定理 2　假定可行集 X 非空，算法产生的 $(\boldsymbol{x}^k,\boldsymbol{\omega}^k,\boldsymbol{\lambda}^k)$ 包含在一个紧集中. 执行算法并在算法中用 0 替换 η，则要么算法在第 k 步迭代终止且 $\boldsymbol{x}^k$ 是问题 NLP 的解，要么对 $\{\boldsymbol{x}^k\}$ 的任一聚点 $\boldsymbol{x}^*$，$\boldsymbol{x}^*$ 是问题 NLP 的解.

证明　当算法在第 k 步迭代终止时，则 $\phi_i(\boldsymbol{x}^k,\lambda_i^k,C_i^k)=0$，$g_i(\boldsymbol{x}^k)\leqslant 0$，$\lambda_i^k\geqslant 0$，$\lambda_i^kg_i(\boldsymbol{x}^k)=0$.

易知对任意的 $c>0$，有 $\phi_i(\boldsymbol{x}^k,\lambda_i^k,c))=0$，且

$$\lambda_i^{k+1}=\phi_i(\boldsymbol{x}^{k+1},\lambda_i^k,C_i^k)+\lambda_i^k=\lambda_i^k\tag{16}$$

又因为 $\|\Phi(\boldsymbol{x}^k,\lambda_i^k,C_i^k))\|_\infty=0$，$\|H(\boldsymbol{x}^k)\|_\infty=0$，则由式(5)，(16)得

$$\nabla S(\boldsymbol{x}^k,\boldsymbol{\omega}^k,\boldsymbol{\lambda}^k,\boldsymbol{C}^k,\boldsymbol{D}^k)=\nabla L(\boldsymbol{x}^k,\boldsymbol{\omega}^{k+1},\boldsymbol{\lambda}^{k+1})=0\tag{17}$$

$\boldsymbol{x}^k$ 是问题 NLP 解. 另一方面，如果算法不在第 k 次迭代终止，对 $\{(\boldsymbol{x}^k,\boldsymbol{\omega}^k,\boldsymbol{\lambda}^k)\}$ 的任一聚点 $(\boldsymbol{x}^*,\boldsymbol{\omega}^*,\boldsymbol{\lambda}^*)$，从定理 1 得出 $\Phi(\boldsymbol{x}^*,\lambda_i^*,\boldsymbol{C}))\|=0$，$\|H(\boldsymbol{x}^*)\|=0$，其中 C 可以是任意正数.

对于任意的 $\bar{\boldsymbol{x}}\in X$，有

$$f(\bar{\boldsymbol{x}})\geqslant f(\boldsymbol{x}^{k+1})+(\boldsymbol{\omega}^k)^{\mathrm{T}}H(\boldsymbol{x}^{k+1})+\sum_{i=1}^{p}D_i^k(h_i(\boldsymbol{x}^{k+1}))^2/2+\sum_{i=1}^{m}[(\phi_i(\boldsymbol{x}^{k+1},\lambda_i^k,C_i^k)+\lambda_i^k)^2-(\lambda_i^k)^2]/(2C_i^k)=f(\boldsymbol{x}^{k+1})+o(1)\tag{18}$$

令 $k\to\infty$，则

$$f(\bar{\boldsymbol{x}})\geqslant f(\boldsymbol{x}^*)$$

显然，$\boldsymbol{x}^*$ 是问题 NLP 的解.

4　子算法的收敛性

4.1　子算法 1 的收敛性

首先讨论子算法 1 的收敛性. 给定 $\omega^j,\lambda^j,C^j,D^j$，设 $\bar{f}(\boldsymbol{x})=S(\boldsymbol{x},\omega^j,\lambda^j,C^j,D^j)$，$F(\boldsymbol{x})=$

$\nabla\bar{f}(\boldsymbol{x})$,序列$\{\boldsymbol{x}^k\}$ 由子算法 1 给出.

给定初始值 $\boldsymbol{x}^k$,$\beta^0\in(0,1)$,$\tau\in(0,1)$,$\theta\in(0,1)$和一列数$\{\bar{\eta}^k\}$,$0<\bar{\eta}^k<1$,$k=0$,取充分小的 ε.

(1) 选择 $\boldsymbol{V}^k\in\partial F^k$,搜索 $\bar{s}^k$ 使得 $\|F^k+\boldsymbol{V}^k\bar{\boldsymbol{s}}^k\|\leqslant\bar{\eta}^k\|F^k\|$.

(2) 如果对于$\boldsymbol{s}^k=\bar{s}^k$,$\eta^k=\bar{\eta}^k$,以下条件(19)和(6)成立,则令 $\boldsymbol{s}^k=\bar{\boldsymbol{s}}^k$,$\boldsymbol{\eta}^k=\bar{\boldsymbol{\eta}}^k$,否则令 $\beta^k=-\operatorname{sgn}((F^k)^{\mathrm{T}}\boldsymbol{V}^k\boldsymbol{F}^k)\min\{1,\|F^k\|\}\beta^0$,$\hat{\eta}^k=\|\boldsymbol{F}^k+\boldsymbol{V}^k\bar{\boldsymbol{s}}^k\|/\|\boldsymbol{F}^k\|$,$\eta^k=1-\tau^i(1-\hat{\eta}^k)$且$\boldsymbol{s}^k=[-\beta^k(1-\eta^k)^{1/2}\cdot F^k+(1-\eta^k)\bar{s}^k]/(1-\hat{\eta}^k)$,其中 i 是满足以下两条件的最小非负整数:

$$\|\boldsymbol{r}^k\|=\|F^k+\boldsymbol{V}^k\boldsymbol{s}^k\|\leqslant\eta^k\|F^k\| \tag{19}$$

$$\|F(\boldsymbol{x}^k+\boldsymbol{s}^k)\leqslant[1-\theta(1-\eta^k)]\|\boldsymbol{F}^k\| \tag{20}$$

(3) 如果$\|\boldsymbol{F}^k\|\leqslant\varepsilon$,算法停止.否则令 $\boldsymbol{x}^{k+1}=\boldsymbol{x}^k+\boldsymbol{s}^k$,转1.

下面引理 4—6 保证子算法 1 是合理的;定理 3—5 表明算法有全局收敛性和超线性收敛性,详细证明参考文献[10-11].

定理 3 假设 $\eta^k<1$ 时执行子算法 1.如果 $\sum_{k=1}^{\infty}(1-\eta^k)$ 是发散的,则$\lim_{k\to\infty}\boldsymbol{F}^k=0$.

条件 1 假设函数 F 是 Lipschitz 连续的.如果对于任意的 $\boldsymbol{y}\in\mathbf{R}^n$ 和任意的 $\boldsymbol{V}_y\in\partial F(\boldsymbol{y})$,下式成立:

$$F(\boldsymbol{y})-F(\boldsymbol{x})=\boldsymbol{V}_y(\boldsymbol{y}-\boldsymbol{x})+o(\|\boldsymbol{y}-\boldsymbol{x}\|)$$

则称 F 在点$\boldsymbol{x}$ 满足条件 1.

条件 2 如果对于任意的$\boldsymbol{V}_x\in\partial F(\boldsymbol{x})$,有

$$|F(\boldsymbol{x})^{\mathrm{T}}\boldsymbol{V}_xF(\boldsymbol{x})|\geqslant 2\|F(\boldsymbol{x})\|^2/M,$$
$$\|\boldsymbol{V}_x\|\leqslant M/2,\ \|\boldsymbol{V}_x^{-1}\|\leqslant M/2 \tag{21}$$

其中 $M>1$ 是一常数,则称 F 在点$\boldsymbol{x}$ 满足条件 2.

$\bar{f}(\boldsymbol{x})=S(\boldsymbol{x},\boldsymbol{\omega}^k,\boldsymbol{\lambda}^k,\boldsymbol{C}^k,\boldsymbol{D}^k)$,$F(\boldsymbol{x})=\nabla\bar{f}(\boldsymbol{x})$在$\boldsymbol{x}$ 点处是凸函数,则 F 在点$\boldsymbol{x}$ 满足条件 2.

引理 4 若 F 在点$\boldsymbol{x}$ 满足条件 2,则对任意的$\boldsymbol{y},\boldsymbol{z}\in B_\delta(\boldsymbol{x})$,任意的$\boldsymbol{V}_y\in\partial F(\boldsymbol{v})$,存在 $\delta>0$ 使得

$$\|\boldsymbol{V}_y\|\leqslant M\quad\|\boldsymbol{V}_y^{-1}\|\leqslant M \tag{22}$$

其中 $M\geqslant1$ 且

$$|F(\boldsymbol{z})^{\mathrm{T}}\boldsymbol{V}_yF(\boldsymbol{z})|\geqslant\|F(\boldsymbol{z})\|^2/M \tag{23}$$

引理 5 给出满足 $F(\boldsymbol{x})\neq0$ 的向量$\boldsymbol{x}$,矩阵$\boldsymbol{V}_y$.假设$\|\boldsymbol{V}_y\|\leqslant M$,$\|\boldsymbol{V}_y^{-1}\|\leqslant M$,$|F(\boldsymbol{x})^{\mathrm{T}}\boldsymbol{V}_yF(\boldsymbol{x})|\geqslant\|F(\boldsymbol{x})\|^2/M$.如果存在 $\bar{s}$ 对于某个 $\hat{\eta}\in(0,1)$满足$\|F(\boldsymbol{x})+\boldsymbol{V}_x\boldsymbol{s}\|=\hat{\eta}\|F(\boldsymbol{x})\|$.令 $\beta=-\operatorname{sgn}(F(\boldsymbol{x})^{\mathrm{T}}\times\boldsymbol{V}_xF(\boldsymbol{x}))\min\{1,\|F(\boldsymbol{x})\|\}\beta_0$,$\beta_0\in(0,1)$是个常数且

$$1>\eta\geqslant\max\left\{\frac{1+\hat{\eta}}{2},\ 1-\left[\frac{|\beta|(1-\hat{\eta})}{8M^3}\right]^2\right\} \tag{24}$$

则向量$\boldsymbol{s}=[-\beta(1-\eta)^{1/2}F(\boldsymbol{x})+(1-\eta)\bar{s}]/(1-\hat{\eta})$满足

$$\|F(\boldsymbol{x})+\boldsymbol{V}_y\boldsymbol{s}\|\leqslant\|F(\boldsymbol{x})\|\left[\eta-\frac{|\beta|(1-\eta)^{1/2}}{M(1-\hat{\eta})}\right] \tag{25}$$

引理 6 给定满足 $F(\boldsymbol{x})\neq0$,$\boldsymbol{V}_x\in\partial F(\boldsymbol{x})$,$\theta\in(0,1)$的$\boldsymbol{x}$.假设 $F(\boldsymbol{x})$在点$\boldsymbol{x}$ 满足条件 1 和条件 2.如果对于某个 $\hat{\eta}\in(0,1)$,存在 $\bar{s}$ 使得$F(\boldsymbol{x})+\boldsymbol{V}_x\bar{s}\|=\hat{\eta}\|F(\boldsymbol{x})\|$,则存在$\lambda\in(0,1)$使得对于任意的 $\eta\in(\lambda,1)$,向量$\boldsymbol{s}=[-\beta(1-\eta)^{1/2}F(\boldsymbol{x})+(1-\eta)\bar{s}]/(1-\hat{\eta})$满足

$$\|F(\boldsymbol{x})+\boldsymbol{V}_x\boldsymbol{s}\|\leqslant\eta\|F(\boldsymbol{x})\|,$$
$$\|F(\boldsymbol{x}+\boldsymbol{s})\|\leqslant[1-\theta(1-\eta)]\|F(\boldsymbol{x})\| \tag{26}$$

在上述$\boldsymbol{s}$ 的表述中,$\beta=-\operatorname{sgn}(F(\boldsymbol{x})^{\mathrm{T}}\boldsymbol{V}_xF(\boldsymbol{x}))\cdot\min\{1,\|F(\boldsymbol{x})\|\}\beta_0$,$\beta_0\in(0,1)$是个常数.

定理 4 在 $\bar{\eta}^k\leqslant\eta_{\max}<1$ 条件下执行子算法 1 得到序列$\{\boldsymbol{x}^k\}$.如果$\boldsymbol{x}^*$是$\{\boldsymbol{x}^k\}$的聚点,F 在点$\boldsymbol{x}^*$满足条件 1 和条件 2,则 $F^*=F(\boldsymbol{x}^*)=0$.

定理 5 在条件 $\bar{\eta}^k\leqslant\eta_{\max}<1$,$\bar{\eta}^k\to0$ 条件下执行子算法 1 产生序列$\{\boldsymbol{x}^k\}$.如果 $\boldsymbol{x}^*$ 是$\{\boldsymbol{x}^k\}$的聚点且 F 在点 $\boldsymbol{x}^*$ 满足条件 1 和条件 2,则 $\boldsymbol{x}^k$ 超线性收敛于$\boldsymbol{x}^*$.

4.2 子算法 2 的收敛性

修正 Broyden 算法为迭代算法.令$\boldsymbol{z}^k=((\boldsymbol{x}^k)^{\mathrm{T}},(\boldsymbol{\omega}^k)^{\mathrm{T}},(\boldsymbol{\lambda}^k)^{\mathrm{T}})^{\mathrm{T}}$,$\boldsymbol{S}^k=S((\boldsymbol{x}^k,\boldsymbol{\omega}^k,\boldsymbol{\lambda}^k))$.给出一个初始点 $\boldsymbol{z}^1$ 和一个初始的正定矩阵 $\boldsymbol{B}^1$,算法生成一个点序列$\{\boldsymbol{z}^k\}$和一个矩阵序列$\{\boldsymbol{B}^k\}$,这两个序列由式(27)和 BFGS(Broyden-Fletcher-Goldfarb-Shanno)修正给出

$$\boldsymbol{z}^{k+1}=\boldsymbol{z}^k+\boldsymbol{s}^k=\boldsymbol{z}^k+\alpha^k\boldsymbol{d}^k \tag{27}$$

其中 $\alpha^k>0$ 是步长因子,$\boldsymbol{d}^k$ 是搜索方向并满足$-\boldsymbol{d}^k=\boldsymbol{H}^k\nabla S^k+\|Q^k\boldsymbol{H}^k\nabla S^k\|R^k\nabla S^k$,其中$\nabla S^k$ 是$f(\boldsymbol{x})$在 $\boldsymbol{z}^k$ 的梯度,$\boldsymbol{H}^k$ 是$\boldsymbol{B}^k$ 的逆.

$\{Q^k\}$和$\{Q^k\}$是两个一致有界的正定矩阵序列或半正定矩阵序列.这些矩阵的所有的特征值包含在$[q,r]$中,$0\leqslant q\leqslant r$,即对于所有的 k,$\boldsymbol{x}\in$

$\mathbf{R}^{n+m+p}$，$\boldsymbol{x}\neq 0$ 有

$$q\|\boldsymbol{x}\|^2 \leqslant \boldsymbol{x}^{\mathrm{T}} Q^k \boldsymbol{x} \leqslant r\|\boldsymbol{x}\|^2$$
$$q\|\boldsymbol{x}\|^2 \leqslant \boldsymbol{x}^{\mathrm{T}} R^k \boldsymbol{x} \leqslant r\|\boldsymbol{x}\|^2 \qquad (28)$$

假设 $f(\boldsymbol{x})\in C^{1,1}$ 和水平集 $S(\boldsymbol{x}_1)=\{\boldsymbol{x}\mid, f(\boldsymbol{x})\leqslant f(\boldsymbol{x}_1)\}$ 有界，得到如下定理，详细证明参考文献[9].

定理 6　算法是全局收敛的.

5　讨论

对于优化约束问题，终止准则为 $\|g\|\leqslant 10^{-5}$. 表 1 中，NIT 为迭代的次数，NS 为计算函数的次数，NG 为计算梯度的次数.

表 1　数值试验结果

Tab.1　Numerical test results

来自文献[2]的问题编号	初始点	NIT	NS	NG	初始点	NIT	NS	NG
227	0.5,0.5	10	19	27	1.0,1.0	9	21	28
227	10.0,10.0	7	15	31	−10.0,−10.0	12	16	21
215	0.5,0.5	8	13	23	1.0,1.0	10	12	32
215	1.5,1.5	6	12	20	2.0,2.0	6	17	24
232	2.0,0.5	5	6	7	4.0,1.0	6	8	11
232	4.0,2.0	5	7	7	6.0,2.0	7	10	11
250	5.0,5.0,5.0	8	14	21	−5.0,−5.0,−5.0	10	14	20
250	10.0,10.0,10.0	9	14	24	−10.0,−10.0,−10.0	12	17	25
264	0,0,0,0	11	17	21	0,0.5,1.5,−0.5	14	19	23
264	0,0.8,1.8,−0.8	12	16	19	1.0,1.0,1.0,1.0	11	29	22

可以使用其他的 NCP 函数来代替增广的拉格朗日函数中的 Fischer-Burmeister NCP 函数，例如，使用如下的 3-1分段线性 NCP 函数[13].

$$\psi(a,b)=\begin{cases} 3a-a^2/b & \text{如果 } b\geqslant a>0\text{，或 } 3b>-a\geqslant 0 \\ 3b-b^2/a & \text{如果 } a>b>0 \text{ 或 } 3a>-b\geqslant 0 \\ 9a+9b & \text{如果 } 0\geqslant a \text{ 且 } -a\geqslant 3b\text{，或 } -3a\leqslant b\leqslant 0 \end{cases} \qquad (29)$$

可以得到和本文同样的结果.

参考文献：

[1] Pillo G D, Grippo L. A new class of augmented Lagrangians in nonlinear[J]. SIAM J Contral Opt, 1979, 17: 616.

[2] Pillo G D, Grippo L. An augmented Lagrangian for inequality constraints in nonlinear programming problems [J]. J Optimization Theorem and Applications, 1982, 36: 495.

[3] Pillo G D. Exact penalty method[M]. Boston: Kluwer Ac Press, 1994.

[4] Pillo G D, Lucidi S. On exact augmented Lagrangian functions in nonlinear programming problems [M]. New York: Plenum Press, 1996.

[5] Pillo G D, Lucidi S. An augmented Lagrangian function with improved exactness properties[J]. SIAM J Optimization, 2001, 12: 376.

[6] 濮定国. 一类增广的拉格朗日乘子函数[J]. 上海铁路科技学院学报，1984, 5: 45.
PU Dingguo. A class of augmented Lagrangian multiplier function [J]. Journal of Shanghai Institute of Railway Technology, 1984, 5: 45.

[7] Fischer A. A special Newton-type optimization method [J]. Optimization, 1992, 24: 269.

[8] Pu D, Gui S, Tian W. A class of revised Broyden algorithms without exact line search [J]. Journal of Computational Mathematics, 2004, 22: 11.

[9] Pu D, Zhou Y, Zhang Z. A QP free feasible method[J]. Journal of Computational Mathematics, 2004, 22: 651.

[10] Pu D, Zhang J. Inexact generalized Newton method for second order C-Differentiable optimization[J]. J of Computational and Applied Mathematics, 1998, 93: 107.

[11] Pu D, Tian W. Globally convergent inexact generalized Newton's method for non-smooth equation[J]. J Comput and Applied Math, 2002, 138: 37.

[12] Schittkowski K. More test examples for nonlinear programming codes[M]. Berlin: Springer-Verlag, 1988.

[13] Pu D, Zhou Y. Piecewise linear NCP function for QP-free feasible method[J]. Applied Mathethematics A J of Chinese Universities, 2006, 21(3): 289.