中国农业大学学报　1999,4(4):10～13
Journal of China Agricultural University

对 Choleski-PCG Newton 算法的一些改进①

钟　萍②　　邓　联
(中国农业大学工程基础科学部)　(中国农业大学电子电力工程学院)

摘　要　对无约束最优化问题提出了一种不精确牛顿算法模型 ACPN(α),是对 Deng N. Y. 和 Wang Z. Z. 文(Can Newton method be surpassed. 见 *Chinese Science Bulletin*, 1998, Vol. 43, No. 20, p. 132～134)中 Choleski-PCG Newton 算法的改进。新算法对于变量个数在 35～186 范围内的无约束问题更有效,并打破了所构造的点列必须恰 Q-2阶收敛的局限,对进一步改进算法有提示作用。

关键词　Choleski 分解; Choleski-PCG Newton 算法; 条件预优共轭梯度法
分类号　O 221.2

Some Improvement on Choleski-PCG Newton Method

Zhong Ping
(College of Applied Engineering Sciences, CAU)
Deng Lian
(College of Electronic and Electric Power Engineering, CAU)

Abstract　An inexact Newton method to the unconstrained optimization problem, algorithm ACPN(α), is proposed as some improvement on Choleski-PCG Newton method given in reference Deng N. Y. and Wang Z. Z. (Can Newton method be surpassed. In: *Chinese Science Bulletion*, 1998, Vol. 43, No. 20, p. 132－134). It's more efficient when variable quantities are between 35 and 186 without the constraints of precisely quadratic convergence.

Key words　Choleski factorization; Choleski-PCG Newton method; preconditioned conjugate iteration

考虑求解无约束最优化问题 min $f(x)$ $x\in \mathbf{R}^n$,其中 $f(x):\mathbf{R}^n \mapsto \mathbf{R}$ 是光滑函数,在使用 Newton 法由当前点 x^k 求下一个迭代点时,是通过解牛顿方程得到增量 s^k 的,每步迭代的工作量包括计算海色阵、梯度值和解牛顿方程。文献[1]中提出了 Choleski-PCG Newton 算法,结合使用 Choleski 分解和条件预优共轭梯度法(PCG)求解牛顿方程,使工作量在 $n\geqslant 55$ 时明显少于单纯使用 Choleski 分解的牛顿法。笔者证明了在 Ostrowski 效率意义[2]下,对于变量个数在 35～186 之间的无约束问题,所建立的算法比文献[1]更有效。

1　算法模型 ACPN(α)

交替使用 Choleski 分解和 PCG 子迭代求解牛顿方程,即 Alternate Choleski-PCG Newton(α)算法模型,简记为 ACPN(α),步骤如下。

收稿日期:1999-01-11
①国家自然科学基金资助项目
②钟　萍,北京清华东路 17 号 中国农业大学(东校区)71 信箱,100083

1)置初始值:选取正参数 $1<\alpha\leqslant 2$,置初始点 x^0,$k=0$;

2)终止检验:若 $\|\nabla f(x^k)\|=0$,则停止;

3)开关检验:若 k 为偶数,则转 4);否则转 5);

4)Choleski 分解步:令 $\boldsymbol{B}^k=\nabla^2 f(x^k)$,用 Choleski 分解 $\nabla^2 f(x^k)=\boldsymbol{L}_k\boldsymbol{D}_k\boldsymbol{L}_k^{\mathrm{T}}$,求解牛顿方程

$$\nabla^2 f(x^k)s=-\nabla f(x^k) \tag{1}$$

得到解 s^k,转 6);

5)PCG 步:令 $\boldsymbol{B}^k=\boldsymbol{B}^{k-1}$,$s^0=0$,取预优阵 $\boldsymbol{M}=(\boldsymbol{B}^k)^{-1}$,用算法 PCG[3]得到 s^k。

6)计算新迭代点:令 $x^{k+1}=x^k+s^k$,$k=k+1$,转 2)。

2　ACPN(α)效率分析

2.1　ACPN(α)的收敛速率

假定问题 $\min f(x)$,$x\in\mathbf{R}^n$ 有解 x^*,且 f 在 x^* 的某个开邻域内满足:A1. $f(x)$ 2 次连续可微;A2. $\nabla^2 f(x)$ Lipschitz 连续;A3. $\nabla^2 f(x^*)$ 是对称正定矩阵;A4. 对任意的 n 维向量 $\boldsymbol{h}\neq 0$,有 $\nabla^3 f(x^*)\boldsymbol{hh}\neq 0$。

定理 1　假设 A1～A3 成立,$1<\alpha\leqslant 2$,则存在 $\delta>0$,$M_1>0$,使得如果 $\|x^0-x^*\|\leqslant\delta$,那么由 ACPN($\alpha$)产生的 $\{x^k\}$ 2 步 2α 阶收敛,即对充分大的 k 有

$$\|x^{k+1}-x^*\|\leqslant M_1\|x^{k-1}-x^*\|^{2\alpha}$$

2.2　估计 ACPN(α)中 PCG 子迭代的次数

在对 ACPN(α)进行效率估计之前,需要估计出由当前点 x^k 在 PCG 步中得到增量 s^k 所需要的 PCG 子迭代的次数。

定理 2　假设 A1～A4 成立,$1<\alpha\leqslant 2$ 且 $\alpha\neq 1.25, 1.50, 1.75, 2.00$。如果 $\|x^0-x^*\|\leqslant\delta$,那么对于充分大的 k,ACPN(α)中 PCG 的迭代次数不超过 $t=\lceil 4(\alpha-1)\rceil$。

证明　以下分 3 步完成该定理的证明。

1)设 x^k 是由 ACPN(α)中的 Choleski 分解步得到的,估计 $[\nabla^2 f(x^k)]_{ij}$ 与 PCG 步中 $(\boldsymbol{B}^k)_{ij}$ 的差。其中 $(\cdot)_{ij}$ 表示矩阵 $(\cdot)$ 的第 i 行第 j 列元素。由 A1～A4 成立可知,若 x^k 是由 x^{k-1} 经 Choleski 分解步得到的,那么当 $\|x^0-x^*\|\leqslant\delta$ 时,存在 $N>0$,使 $N\|x^{k-1}-x^*\|^2\leqslant\|x^k-x^*\|$[2]。由 A2 成立可知当 k 充分大时存在 $L>0$ 使 $|[\nabla^2 f(x^k)]_{ij}-(\boldsymbol{B}^k)_{ij}|=|[\nabla^2 f(x^k)]_{ij}-[\nabla^2 f(x^{k-1})]_{ij}|\leqslant L\|x^k-x^{k-1}\|\leqslant L(1+1/N)\|x^k-x^*\|^{1/2}$;又因为 $\|\nabla f(x^k)\|=\|\nabla f(x^k)-\nabla f(x^*)\|\geqslant(1/2)\lambda_{\min}\|x^k-x^*\|$,其中 $\lambda_{\min}$ 为 $\nabla^2 f(x^*)$ 的最小特征值:所以 $[\nabla^2 f(x^k)]_{ij}-(\boldsymbol{B}^k)_{ij}|\leqslant M_2\|\nabla f(x^k)\|^{1/2}$,其中 $M_2=(L/4)(1+1/N)\lambda_{\min}^{1/2}>0$。

2)令 $\boldsymbol{A}^k=(\boldsymbol{B}^k)^{-1}\nabla^2 f(x^k)$,估计 $\boldsymbol{A}^k$ 的2-范数意义下的条件数 $\kappa(\boldsymbol{A}^k)$,即要证明

$$\kappa(\boldsymbol{A}^k)\leqslant(1+\beta^k)(1-\beta^k)^{-1} \tag{2}$$

其中 $\beta^k=nM_2\|(\boldsymbol{B}^k)^{-1}\|_\infty\|\nabla f(x^k)\|^{1/2}$。

事实上 $\boldsymbol{A}^k=(\boldsymbol{B}^k)^{-1}\nabla^2 f(x^k)=\boldsymbol{I}+(\boldsymbol{B}^k)^{-1}[\nabla^2 f(x^k)-\boldsymbol{B}^k]$,$\|(\boldsymbol{B}^k)^{-1}[\nabla^2 f(x^k)-\boldsymbol{B}^k]\|_\infty\leqslant nM_2\|(\boldsymbol{B}^k)^{-1}\|_\infty\|\nabla f(x^k)\|^{1/2}=\beta^k$。由盖尔圆定理可知,$\boldsymbol{A}^k$ 的任一特征值 λ 满足 $1-\beta^k\leqslant\lambda\leqslant 1+\beta^k$,对充分大的 k 又有 $\beta^k\leqslant 1/2$,所以式(2)成立。令 $\mathrm{K}(\boldsymbol{A}^k)$ 是 $\boldsymbol{A}^k$ 的 K-条件数,将证明对充分大的 k 有 $\mathrm{K}(\boldsymbol{A}^k)\leqslant(1+\bar{\beta}^k)(1-\bar{\beta}^k)^{-1}$,其中 $\bar{\beta}^k=(1+n)\beta^k$。事实上,对充分大的 k 有 $\bar{\beta}^k\leqslant 1/2$。由

文献[3]定理 13.5(a)和式(2)可知，对于充分大的 k 有 $\mathrm{K}(\boldsymbol{A}^k)\leqslant\kappa^n(\boldsymbol{A}^k)\leqslant[(1+\beta^k)(1-\beta^k)^{-1}]^n\leqslant[1+(1+n)\beta^k][1-(1+n)\beta^k]^{-1}=(1+\bar{\beta}^k)(1-\bar{\beta}^k)^{-1}$。

3)现需证明当 PCG 的迭代次数达到 $t=\lceil 4(\alpha-1)\rceil$ 时，必定满足 PCG 子迭代的终止准则。当 k 充分大时，由文献[3]定理 13.11(b)知 $\|r^t\|^2_{(B^k)^{-1}}\leqslant[\mathrm{K}(\boldsymbol{A}^k)^{1/t}-1]^t\|r^0\|^2_{(B^k)^{-1}}$，其中 $\|\cdot\|_B$ 是 $\boldsymbol{B}$-范数，r^t 是算法 PCG 迭代 t 次后的残差。而 $\mathrm{K}(\boldsymbol{A}^k)^{1/t}-1\leqslant[(1+\bar{\beta}^k)(1-\bar{\beta}^k)^{-1}]^{1/t}-1\leqslant(1+\bar{\beta}^k)(1-\bar{\beta}^k)^{-1}-1\leqslant 4\bar{\beta}^k=4(1+n)nM_2\|(\boldsymbol{B}^k)^{-1}\|_\infty\cdot\|\nabla f(x^k)\|^{1/2}$；又因为 $\|r^t\|_{(B^k)^{-1}}=\|(\boldsymbol{B}^k)^{-1/2}r^t\|$：所以，对于充分大的 k，有 $\|r^t\|\leqslant\|(\boldsymbol{B}^k)^{1/2}\|\|(\boldsymbol{B}^k)^{-1/2}r^t\|\leqslant 4\|[\nabla^2 f(x^*)]^{1/2}\|\|[\nabla^2 f(x^*)]^{-1/2}\|\cdot[8n(1+n)M_2\|[\nabla^2 f(x^*)]^{-1}\|_\infty]^{t/2}\cdot\|\nabla f(x^k)\|^{1+t/4-\alpha}\|\nabla f(x^k)\|^\alpha$。当 $t=\lceil 4(\alpha-1)\rceil$ 且 $\alpha\neq 1.25$，1.50，1.75，2.00 时即有 $1+t/4-\alpha>0$，从而可以使当 k 充分大时有 $\|r^t\|\leqslant\|\nabla f(x^k)\|^\alpha$，即 PCG 子迭代的终止准则成立。

2.3 估计 ACPN(α)的效率

用 $e_a(\alpha)$ 表示 ACPN(α)的效率，由定理 1 和定理 2 可知，在 Ostrowski 效率[2]意义下有 $e_a(\alpha)\geqslant(\ln 2\alpha)(W_C+tW_{CG}+2d)^{-1}=\eta(\alpha,n,d)$。其中：$W_C$ 是用 Choleski 分解技术求解式(1)的工作量(乘除法的次数和)，$W_C=n^3/6+3n^2/2-2n/3$；W_{CG} 是 1 次 PCG 子迭代的工作量(乘除法的次数和)，$W_{CG}=2n^2+6n+2$；d 为求某点的海色阵和梯度值的工作量；t 是由定理 2 给出的至多需要调用 PCG 子迭代的次数。

考虑问题 $\max\,\eta(\alpha,n,d)$，$1<\alpha\leqslant 2$。令 $\Delta_1=(1.00,1.25)$，$\Delta_2=(1.25,1.50)$，$\Delta_3=(1.50,1.75)$，$\Delta_4=(1.75,2.00)$，对于固定的 n 和 d 值，$\eta(\alpha,n,d)$ 是 α 在 $\Delta_i(i=1,2,3,4)$ 上的严格单调增函数。事实上，对于任意给定的 $\alpha\in\Delta_i$，由定理 2 确定的 t 值相同，由于 $\ln 2\alpha$ 在该区间上是单调增函数，所以 $\eta(\alpha,n,d)$ 也是该区间上的单调增函数。又因为要求该问题的解满足 $1+t/4-\alpha>0$，所以 α 可能的取值为 $\alpha_1=1.25-\varepsilon$，$\alpha_2=1.50-\varepsilon$，$\alpha_3=1.75-\varepsilon$，$\alpha_4=2.00-\varepsilon$。其中 ε 为任意给定的充分小的正数。相应的 PCG 的迭代次数分别为 $t_1=1$，$t_2=2$，$t_3=3$，$t_4=4$。

3 结　论

本文结论可归结为表 1 和表 2。

表 1　不同条件下 Choleski-PCG Newton 算法与 ACPN(α)效率下界的比较

n	d	选用算法
$1\leqslant n\leqslant 31$	—	Newton
$32\leqslant n\leqslant 35$	$d<B_1(n)$	ACPN(α_1)
	$d>B_1(n)$	Newton
$36\leqslant n\leqslant 37$	$d<B_2(n)$	ACPN(α_2)
	$B_2(n)<d<B_1(n)$	ACPN(α_1)
	$d>B_1(n)$	Newton
$38\leqslant n\leqslant 39$	$d<B_1(n)$	ACPN(α_1)
	$B_1(n)<d<B_2(n)$	ACPN(α_2)
	$d>B_2(n)$	Newton
$40\leqslant n\leqslant 41$	$d<B_3(n)$	ACPN(α_3)
	$B_3(n)<d<B_1(n)$	ACPN(α_1)
	$B_1(n)<d<B_2(n)$	ACPN(α_2)
	$d>B_2(n)$	Newton

n	d	选用算法
$n=42$	$d<A_1(n)$	ACPN(α_1)
	$A_1(n)<d<A_2(n)$	ACPN(α_2)
	$A_2(n)<d<A_3(n)$	ACPN(α_3)
	$d>A_3(n)$	ACPN(α_4)
$43\leqslant n\leqslant 55$	$d<A_2(n)$	ACPN(α_2)
	$A_2(n)<d<A_3(n)$	ACPN(α_3)
	$d>A_3(n)$	ACPN(α_4)
$56\leqslant n\leqslant 70$	$d<A_3(n)$	ACPN(α_3)
	$d>A_3(n)$	ACPN(α_4)
$71\leqslant n\leqslant 186$	—	ACPN(α_4)
$n\geqslant 187$	—	Choleski-PCG Newton

表中：$A_1(n)=(1/2)(\ln 2.5/\ln 1.2-1)W_{CG}-(1/2)W_C$；

$A_2(n)=(1/2)[\ln 3/\ln(3.5/3)-2]W_{CG}-(1/2)W_C$；

$A_3(n)=(1/2)[\ln 3.5/\ln(4/3.5)-3]W_{CG}-(1/2)W_C$；

$B_1(n)=(\ln 1.25/\ln 1.6)W_C-(\ln 2/\ln 1.6)W_{CG}$；

$B_2(n)=[\ln 1.5/\ln(4/3)]W_C-[\ln 4/\ln(4/3)]W_{CG}$；

$B_3(n)=[\ln 1.75/\ln(4/3.5)]W_C-[\ln 8/\ln(4/3.5)]W_{CG}$。

表 2　不同条件下算法 ACPN(α)与算法 Choleski-PCG Newton 的效率下界的比值 $r(d=0)$

n	32	35	39	42	54	60	70	186
r	0.999	0.981	0.961	0.948	0.885	0.895	0.911	0.999

参 考 文 献

1 Deng N Y and Wang Z Z. Can Newton method be surpassed. Chinese Science Bulletin, 1998, 43(20): 132～134

2 李庆扬，莫孜中，祁力群. 非线性方程组的数值解法. 北京：科学出版社，1992. 35～37

3 Kelley C T. Iterative methods for linear and nonlinear equations. Philadelphia: SIAM, 1995. 22～23