中国农业大学学报 2001, 6(2): 24～28
Journal of China Agricultural University

应用自动微分技术的CF-PCG方法及其效率分析①

张海斌②
(中国农业大学工程基础科学部)

摘　要　CF-PCG算法是牛顿法和预优共轭梯度法结合起来解牛顿方程的一种非精确牛顿法。笔者将自动微分技术应用到该算法中,并证明应用自动微分技术的CF-PCG方法具有更高的效率。

关键词　牛顿法; 预优共轭梯度法; 自动微分
中图分类号　O 221.2

CF-PCG Method Using Automatic Differentiation and Its Efficience Analysis

Zhang Haibin
(College of Applied Engineering Sciences, CAU)

Abstract　Newton-PCG method is an inexact Newton-like method. It is an organic combination of Newton's method and preconditioned conjugate gradient method. Automatic differentiation method are applied in this algorithm, it has been shown that CF-PCG method appling automatic differentiation techniques is more efficient than the original CF-PCG method using symbolic differentiation method.

Key words　Newton's method; preconditioned conjugate gradient method; automatic differentiation

牛顿法是非线性优化中最基础的算法之一,许多非精确牛顿法的出现从不同角度改进了牛顿法。邓乃扬在文献[1],[2]中提出了一类既保持牛顿法的收敛速度又降低计算成本的改进的非精确牛顿法。该算法的基本结构是先进行一个精确牛顿步,即利用Cholesky分解去解牛顿方程,简称CF步;接着进行p个PCG步,即进行p次预优共轭梯度算法解牛顿方程。自动微分技术可降低计算成本,因而可用来改进算法。

1　应用自动微分技术的CF-PCG方法

1.1　关于自动微分

考虑函数$f(x)$：$\mathbf{R}^n\to\mathbf{R}$,其梯度和Hessian矩阵分别用$g(x)$,$G(x)$表示,种子向量用$\dot{x}$表示。

算法AD　给定初始点x和种子向量$\dot{x}$,下面步骤可以得到$f(x)$,$g(x)$,$G(x)\dot{x}$

第1步　使用链式规则得到$f(x)$;

收稿日期: 2000-07-17
①国家自然科学基金资助项目
②张海斌,北京清华东路17号 中国农业大学(东校区)71号信箱, 100083

第 2 步　利用逆向规则计算函数梯度 $g(x)$;

第 3 步　令 $F(x)=\{f(x), g(x)\}: \mathbf{R}^n \to \mathbf{R}^{n+1}$, 对 $F(x)$ 再次使用顺向规则便得到 $G(x)\dot{x}$。

说明: 可以在第 2 步与第 3 步之间设置 2 个开关, 当仅需要 $g(x)$ 时, 输出 $g(x)$ 而不继续第 3 步; 当需计算 $G(x)$ 时, 可以分别取 $\dot{x}=e_1, e_2, \cdots, e_n$, 并输出 $G(x)$, 其中 $e_1, e_2, \cdots, e_n$ 为单位矩阵 I_n 的行向量[3]。

定理 1　若通过算法 AD 计算函数 $f(x)$ 的梯度 $g(x)$ 及 Hessian 矩阵 $G(x)$, 则

$$Q\{f(x), g(x), G(x)\dot{x}\} \le 10Q\{f(x)\}$$

$$Q\{f(x), g(x), G(x)\} \le (4+6n)Q\{f(x)\}$$

其中 Q 表示相应的计算成本。

1.2　关于 CF-PCG 算法

考虑无约束优化问题

$$\min_{x \in \mathbf{R}^n} f(x) \tag{1}$$

其中 $x \in \mathbf{R}^n$; $f: \mathbf{R}^n \to \mathbf{R}$ 是一个连续可微函数, 且满足假设:

A1　问题(1)有一个最优解 x^*, 且 Hessian 矩阵 $G(x^*)$ 对称正定;

A2　存在 $\rho > 0$, Hessian 矩阵 $G(z)$ 是 Lipschits 连续的。

用算法 PCG(C, A, b, l, e) 解线性方程组

$$As=b$$

其中: C 是预优阵, l 是最大子迭代次数, e 是终止条件指数。

算法 PCG(C, A, b, l, e)

1) 数据初始化。置 $s_0=0, r_0=b, i=1$。

2) 终止试验。如果

$$\|r_{i-1}\| \le \|b\|^{1+e} \text{或 } i-1=l$$

转步 4)。

3) 子迭代。a. 置 $z=Cr_{i-1}$, $t_{i-1}=z^{\mathrm{T}}r_{i-1}$; b. 若 $i=1$, 则 $\beta=0, q=z$; 否则 $\beta=t_{i-1}/t_{i-2}, q=z+\beta q$; c. 置 $s_i=s_{i-1}+\alpha q$, 其中 $\alpha=t_{i-1}/(q^{\mathrm{T}}\omega)$, $\omega=Aq$; d. 置 $r_i=r_{i-1}-\alpha\omega$。

4) 置 $i=i+1$, 转步 2)。

5) 置 $\tilde{s}=s_{i-1}$。

基于算法 PCG(C, A, b, l, e) 所建立起来的算法模型 CFPCG$(\cdot)$ 包含参数 p 和 σ, 其中 p 是一个 CF 步后跟的 PCG 步数, σ 是 p 个 PCG 步总的子迭代次数。本文中只考虑其中之一, 参数 p 和 σ 是如下一维最优化问题的解。

$$\max_{\sigma \in z} v(\bar{p}(\sigma), \sigma) = \frac{\ln(2+\sigma)}{(4+6n+4\bar{p}(\sigma)+6\sigma)Q_f + Q_D + \sigma Q_I} \tag{2}$$

其中: $\sigma \in N$, N 表示全体非负整数集合; $\bar{p}(\sigma) = \lceil \ln(2+\sigma)/\ln 2 - 1 \rceil$, 表示取不小于 $\ln(2+\sigma)/\ln 2-1$ 的最小整数; Q_f 表示函数 $f(x)$ 的计算量; Q_D 和 Q_I 分别满足

$$Q_D = 1/6n^3 + 3/2n^2 - 2/3n, \quad Q_I = n^2 + 6n + 2$$

算法 CFPCG(σ)

初始步。选取初始点 $x^0 \in \mathbf{R}^n$, 置 $k=0$, 计算 $p=\lceil \ln(2+\sigma)/\ln 2-1 \rceil$。如果 $p>1$, 置 $l_1=2^1$,

$l_2 = 2^2, \cdots, l_{p-1} = 2^{p-1}, l_p = \sigma - 2^p + 2$。

第 1 步　终止试验。若 $\| g(x^k) \| = 0$，结束；

第 2 步　转换试验。若 k 能被 $p+1$ 整除，转第 3 步；否则，转第 4 步；

第 3 步　精确牛顿步(CF)。解牛顿方程

$$G(x^k)s = -g(x^k) \tag{3}$$

利用 Cholesky 分解 $G(x^k) = L_k D_k L_k^T$，置 $m = 0$，转第 5 步。

第 4 步　非精确牛顿步(PCG)。对牛顿方程求近似解 s^k。置 $B^k = B^{k-1}$，$m = m+1$，执行 PCG $((B^k)^{-1}, G(x^k), g(x^k), l_m, 1 + l_m/2^m)$。置 $s^k = \tilde{s}$。

第 5 步　置 $x^{k+1} = x^k + s^k$，$k = k+1$，转第一步。

2　效率分析

算法 CFPCG(σ) 在传统微分方法下的收敛性分析在文献[1]和[2]中已经给出，应用自动微分技术的算法 CFPCG(σ) 也有同样的结论。解一维优化问题式(2)是为了得到最优的效率下界 $v(\sigma^*)$。用 Q_f 表示函数 $f(x)$ 的计算量，Q_g 表示函数梯度计算量，并由定理 1 得到 Hessian 矩阵和梯度计算量 $Q_{Hg} = (4+6n)Q_f$，单独梯度 $Q_g = 4Q_f$。Hessian 矩阵与向量的乘积为 $10Q_f$，若 p 个 PCG 步共包含 σ 次子迭代，则 p 个 PCG 步中 Hessian 矩阵与向量的乘积(包含梯度)的计算量为 $(4+6l)Q_f$，其中 Q_D 为所需的工作量，Q_I 是 PCG 子迭代的算术计算量。

自动微分下，$(4+6n+4p+6\sigma)Q_f + Q_D + \sigma Q_I$ 实际上是算法 CF-PCG(σ) 执行一段(1CF+pPCG)的计算量，而在符号微分下执行一段的计算量为 $1/2(n^2+3n)(p+1)Q_f + Q_D + \sigma(Q_I + n^2)$，因此，应用自动微分技术可使算法的效率提高。

定义 1　相对效率：设算法 CFPCG(σ) 的最优效率下界为 $v(\sigma^*)$（下称为效率），$u \overset{\text{def}}{=} \ln 2/Q_{CF}$ 表示牛顿法的效率，其中 Q_{CF} 为一步牛顿法所需的工作量，则算法 CF-PCG 相对于牛顿法的相对效率 η 为

$$\eta = \frac{v(\sigma^*)}{u} \tag{4}$$

定理 2　假定在符号微分下，$Q_g = nQ_f$，$Q_H = 1/2(n^2+3n)Q_f$，η 由定义 1 给出，$n \geq 63$，则

$$\eta_{AD} > \eta_{SD}$$

为证明定理 2，先给出以下引理。

引理 1　算法 CFPCG(σ) 的最优化问题式(2)的全局解 σ^* 与 p^* 满足

$$\sigma^*/p^* \leq n/6 \tag{5}$$

并且相应于符号微分式(5)成立。

证明　p^*, σ^* 是优化问题式(2)的全局最优解，且 $\sigma^* \in [2^{p^*} - 1, 2^{p^*+1} - 2]$，可分以下 2 种情况进行讨论。

1) 当 $\sigma^* \in (2^{p^*} - 1, 2^{p^*+1} - 2]$ 时，有

$$v(p^*, \sigma^* - 1) \leq v(p^*, \sigma^*) \tag{6}$$

将式(6)代入式(2)得

$$\frac{\ln(2+\sigma^*)}{\ln[(2+\sigma^*)/(1+\sigma^*)]} - \sigma^* < \frac{Q_D + (4+6n)Q_f}{Q_I + 6Q_f} + \frac{4p^* Q_f}{Q_I + 6Q_f}$$

又据

$$\frac{Q_D + (4+6n)Q_f}{Q_I + 6Q_f} + \frac{4p^* Q_f}{Q_I + 6Q_f} < n + \frac{2}{3}p^*$$

得

$$\frac{\ln(2+\sigma^*)}{\ln[(2+\sigma^*)/(1+\sigma^*)]} - \sigma^* < n + \frac{2}{3}p^*$$

令

$$j(x) = (2+x)[\ln(2+x) - \ln(1+x)]$$

由 $x > 0$ 时, $j(x) < 0$, $j(0) = 2\ln 2$ 及 $j(+\infty) = 1$ 得

$$(2+\sigma^*)\ln\frac{2+\sigma^*}{1+\sigma^*} \leq 2\ln 2$$

从而

$$\frac{\ln(2+\sigma^*)}{\ln[(2+\sigma^*)/(1+\sigma^*)]} - \sigma^* > \frac{(2+\sigma^*)\ln(2+\sigma^*)}{2\ln 2} - \sigma^*$$

故

$$0.7(2+\sigma^*)\ln(2+\sigma^*) - \sigma^* < n + \frac{2}{3}p^*$$

由于 $\sigma^* > 52$ 时, $\ln(2+\sigma^*) > 4$, $p^* \geq 5$ 及 $p^* < 2\ln(2+\sigma^*)$, 从而 $\sigma^* < 5n/9$; 所以当 $\sigma^* > 52$ 时有 $\sigma^*/p^* < n/6$。

又当 $\sigma^* \leq 52$ 时, 由于 $\max\{\sigma^*/p^* \mid \sigma^* = 1, 2, \cdots, 52\} \leq 10.4$, 故当 $n \geq 63$ 时, $\sigma^*/p^* \leq 10.4 < n/6$

2) 当 $\sigma^* = 2^{p^*} - 1$ 时, 有

$$v(p^* - 1, \sigma^* - 1) \leq v(p^*, \sigma^*) \tag{7}$$

将式(7)代入式(2)中有

$$\frac{\ln(2+\sigma^*)}{\ln[(2+\sigma^*)/(1+\sigma^*)]} - \sigma^* < \frac{Q_D + (4+6n)Q_f}{Q_I + 10Q_f} + \frac{4(p^* - \sigma^*)Q_f}{Q_I + 10Q_f} < \frac{2}{3}n$$

以下证明同 1)。

对于符号微分时的情形, 证明类似。

引理 2　设 n, l 均为自然数且 $1 \leq l \leq n/6$, $Q_D = 1/6n^3 + 3/2n^2 - 2/3n$, Q_f 是 $f(x)$ 的计算量, 则下列不等式成立:

$$\frac{1/2(n^2+3n)Q_f + Q_D}{(6n+4)Q_f + Q_D} < \frac{1/2(n^2+3n)Q_f + l(2n^2+6n+2)}{(6l+4)Q_f + l(n^2+6n+2)} \tag{8}$$

证明　令

$$\Phi(x) = \frac{1/2(n^2+3n)x + l(2n^2+6n+2)}{(6l+4)x + l(n^2+6n+2)} - \frac{1/2(n^2+3n)x + Q_D}{(6n+4)x + Q_D}$$

易证 $\Phi(0) > 0$, $\Phi(+\infty) > 0$, 如果由 $\Phi'(x) = 0$ 求得的解 $x^* < 0$, 则说明 $\Phi(x)$ 在 $(0, +\infty)$ 无极值点, 故式(8)成立。

由 $\Phi'(x^*) = 0$ 得

$$\{6n + 4 - (6l+4)[nt/(6l)]^{1/2}\}x^* = l(n^2+6n+2)[nt/(6l)]^{1/2} - Q_D \tag{9}$$

其中

$$t \overset{\text{def}}{=} \frac{n^4 - 93n^2 - 36n + 32}{n^4 + 9n^3 - 24ln^2 + 4n^2 - 72ln - 42n - 24l - 16}$$

则由 $l < n/6$ 得 $t \le 1$, $(6n+4)(6l+4) > (n+1)(l+1) > n/(6l) \ge [nt/(6l)]^{1/2}$。故

$$6n + 4 - (6l+4)[nt/(6l)]^{1/2} > 0 \tag{10}$$

由 $t \le 1$ 及 $n \ge 6$ 得 $l(n^2 + 6n + 2)[nt/(6l)]^{1/2} - Q_D < 0$, 结合式(9)和(10)得 $x_l^* < 0$。

引理3 对由一维最优化问题式(2)得到的算法 CFPCG(σ), 由于 $p = \lceil \ln(2+\sigma)/\ln 2 - 1 \rceil > 0$, 故

$$\frac{Q_{SD}^{C}}{Q_{AD}^{C}} < \frac{Q_{SD}^{P}}{Q_{SD}^{P}} \tag{11}$$

式(11)中 Q 表示计算量, 下标表示微分方法, 上标表示所采用的迭代方法。

证明 引理3实际上给出的是 $p = 1$ 时的情形, $p > 1$ 时可类似地证明, 不过其中的 l 换为 σ^*/p^*, 且由引理1知 $\sigma^*/p^* < n/6$ 满足条件, 故引理3结论成立。

定理2的证明 由引理3得

$$\frac{Q_{SD}^{C}}{Q_{AD}^{C}} < \frac{Q_{AD}^{C} + Q_{SD}^{P}}{Q_{AD}^{C} + Q_{AD}^{P}} = \frac{Q_{SD}^{CP}}{Q_{AD}^{CP}} \tag{12}$$

令 $(p_{SD}^*, \sigma_{SD}^*)$, $(p_{AD}^*, \sigma_{AD}^*)$ 分别为 CFPCG(σ) 在符号微分和自动微分下的一维优化问题的解, 由引理1, 它们均满足 $\sigma^*/p^* < n/6$。

由此, 由式(12)得

$$\frac{\ln(2+\sigma_{SD}^*)/Q_{SD}^{CP}(\sigma_{SD}^*)}{\ln 2/Q_{SD}^{C}} < \frac{\ln(2+\sigma_{SD}^*)/Q_{AD}^{CP}(\sigma_{SD}^*)}{\ln 2/Q_{AD}^{C}} \le \frac{\ln(2+\sigma_{AD}^*)/Q_{AD}^{CP}(\sigma_{AD}^*)}{\ln 2/Q_{AD}^{C}}$$

记 $\eta_{AD} = \eta_{AD}(p_{AD}^*, \sigma_{AD}^*)$, $\eta_{SD} = \eta_{SD}(p_{SD}^*, \sigma_{SD}^*)$, 则

$$\frac{\eta_{AD}}{\eta_{SD}} \ge \frac{\eta_{AD}(p_{AD}^*, \sigma_{AD}^*)}{\eta_{SD}(p_{SD}^*, \sigma_{SD}^*)} > 1$$

证毕。

本研究是在邓乃扬教授的指导下完成的, 特此致谢。

参考文献

1 Deng Naiyang, Wang Zhaozhi. Theoretical efficiency of an inexact Newton method. Journal of Optimization Theory and Applications, 2000, 105(1): 97～112

2 Deng Naiyang, Wang Zhaozhi. Can Newton method be surpassed. Chiness Science Bulletin, 1999, 43: 132～134

3 Griewank A. Evaluating Derivatives Principles and Techniques of Algorithmic Differentiation. Philadephia: SIAM, 2000. 39～86