DOI:10.13656/j.cnki.gxkx.1995.02.002
广西科学 Guangxi Sciences 2 (2): 6～9 1995

一个既约差商算法及其整体收敛性*

A Reduced Difference Coefficient Algorithm and Its Global Convergence

简金宝
Jian Jinbao

(广西大学数学与信息科学系 南宁市西乡塘路10号 530004)
(Dept. of Math. and Information Science, Guangxi University, 10 Xixiangtang Road, Nanning, 530004)

摘要 利用差商代替难以计算的精确导数，结合既约梯度法的思想建立新的算法；在目标函数一致凸的条件下证明了既约差商法的整体收敛性.

关键词 约束最优化 既约差商 差商 整体收敛性 一致凸

Abstract The difference coefficient was used to replace the exact derivative which is difficult to be computed, and a new algorithm was presented by using the idea of reduced gradient method. The reduced difference coefficient algorithm was shown to possess global convergence if the objective function is uniformly convex.

Key words constrained optimization, reduced difference coefficient, difference coefficient, global convergence, uniformly convex

自从 Wolfe[1] (1962) 提出既约梯度法以来，经过多年的实践表明该类方法的效果明显优于其它方法[2]. 该类方法在理论上也得到了广泛地研究，形成各种类型的既约梯度法. 但绝大多数可微优化算法的实现均依赖于导数的计算，而多数实际最优化的目标函数的精确导数难以计算，这给算法的应用带来了很大困难.

Dixon[3]和赵小平[4]提出了用差商代替导数的无约束优化变尺度法. 在此，我们借助既约梯度法降维的思想，以差商代替目标函数的精确导数，构造了线性约束最优化的一个既约差商法. 由于方法不需计算任何精确导数，故而具有较小的计算量和广泛的实用性.

1 既约差商算法

我们讨论如下非线性优化问题

$$(P)\qquad \begin{aligned} &\min f(x)\\ &s.t.\quad Ax=b\\ &\qquad\quad x\geqslant 0\end{aligned}$$

其中 $A\in E^{m\times n}, b\in E^m, x\in E^n, f(x): E^n\to E^1$.

对 (P) 作如下假设

(a) 可行集 $R=\{x\in E^n: Ax=b, x\geqslant 0\}$ 是非退化的，即任意基可行解至少有 m 个正分量.

(b) 目标函数 $f(x)$ 二阶连续可微，且存在常数 $M_1, M_2, 0<M_1<M_2$，使得 $\forall\, x,y\in E^n$ 有

$$M_1\|y\|^2\leqslant y^T\nabla^2 f(x)y\leqslant M_2\|y\|^2 \qquad (1)$$

由(1)知 $f(x)$ 为一致凸，从而 (P) 的最优解唯一，设 (P) 的最优解为 x^*.

设 a_j 为 A 的第 j 个列向量，$y\in E^n$. 对于子集 $I\subset\{1,\cdots,n\}$，记

$A_I=(a_j, j\in I)$, $\quad y_I=(y_j, j\in I)$.

如果 $|I|=m$，A_I 为非奇异，称 I 为 (P) 的基.

对于基 I，记 $J=\{1,\cdots,n\}\backslash I$. 则

$$Ax=b\Longleftrightarrow x_I=A_I^{-1}b-A_I^{-1}A_Jx_J \qquad (2)$$

$$\begin{aligned} f(x)&=f(x_I,x_J)\\ &=f(A_I^{-1}b-A_I^{-1}A_Jx_J,x_J)\triangleq \bar f(x_J)\end{aligned} \qquad (3)$$

$$\nabla\bar f(x_J)=\nabla_J f(x)-(A_I^{-1}A_J)^T\nabla_I f(x) \qquad (4)$$

$r(x)\triangleq\nabla\bar f(x_J)$ 称为 $f(x)$ 的既约梯度.

1994-04-08 收稿。
1994-08-03 修回。
*广西大学青年科学基金资助项目。

对于 $h>0$，定义

$$s_i = s_i(x,h) = \frac{f(x+he_i)-f(x)}{h} \tag{5}$$

$$s=(s_1,\cdots,s_n)^T \in E^n.$$

其中 e_i 为 E^n 中第 i 个分量为 1，其余分量为零的向量，$i=1,\cdots,n$.

称 s 为 $f(x)$ 以 h 为步长的差商，简称为 $f(x)$ 的差商；称 $\bar{r}(x)$ 为 $f(x)$ 的既约差商，$\bar{r}(x)$ 定义为

$$\begin{aligned}\bar{r}(x) &= \bar{r}(x,h)\\ &= s_J(x,h)-(A_I^{-1}A_J)^T s_I(x,h)\\ &= s_J-(A_I^{-1}A_J)^T s_I \end{aligned}\tag{7}$$

记 $\bar{f}(x_J)$ 的差商为 $\bar{s}=(\bar{s}_j, j\in J)$. 下面给出 $\nabla f(x), s, r(x), \bar{r}(x), \bar{s}$ 之间的关系

定理 1 在假设(b)下，有

1) $$\nabla f(x)+\frac{h}{2}M_1 e \leqslant s \leqslant \nabla f(x)+\frac{1}{2}hM_2 e \tag{8}$$

$e=(1,\cdots,1)^T\in E^n$.

2) $$r(x)+\frac{1}{2}hM_1(1+\xi)\bar{e} \leqslant \bar{s} \leqslant r(x)+\frac{1}{2}hM_2(1+\eta)\bar{e},\ \bar{e}=\bar{e}_J \tag{9}$$

$\xi=\min\{\|A_I^{-1}a_j\|^2, j\in J\}$,

$\eta=\max\{\|A_I^{-1}a_j\|^2, j\in J\}$.

3) $$\bar{s}+\frac{h}{2}[M_1-M_2(1+\eta+\sqrt{m\eta})]\bar{e} \leqslant \bar{r}(x) \leqslant \bar{s}+\frac{h}{2}[M_2(1+\sqrt{m\eta})-M_1(1+\xi)]\bar{e} \tag{10}$$

证明 1) 由中值定理有

$$s_j=\frac{f(x+he_j)-f(x)}{h} = \nabla f(x)^T e_j+\frac{1}{2}he_j^T\nabla^2 f(y_j)e_j$$

记，$c_j=e_j^T\nabla^2 f(y_j)e_j$，则有

$$s=\nabla f(x)+\frac{1}{2}hC,\quad M_1\leqslant c_j\leqslant M_2 \tag{11}$$

于是 1) 成立.

2) $$\bar{s}_j=\frac{\bar{f}(x_J+h\bar{e}_j)-\bar{f}(x_J)}{h} = \frac{f(x_I-A_I^{-1}A_Jh\bar{e}_j, x_J+h\bar{e}_j)-f(x_I,x_J)}{h}$$

利用中值定理及(b)易推知

$$h\bar{s}_j=\nabla_I f(x)^T(-hA_I^{-1}A_J\bar{e}_j)+h\nabla_J f(x)^T\bar{e}_j + \frac{1}{2}h^2\begin{pmatrix}-A_I^{-1}A_J\bar{e}_j\\ \bar{e}_j\end{pmatrix}^T\nabla^2 f(z_j)\begin{pmatrix}-A_I^{-1}A_J\bar{e}_j\\ \bar{e}_j\end{pmatrix}$$

其中 $\bar{e}_j=(e_j)_J=(e_{ji}, i\in J)$，$e_{ji}$ 为 e_j 第 i 个分量

$$M_1(1+\|A_I^{-1}a_j\|^2) \leqslant \begin{pmatrix}-A_I^{-1}A_J\bar{e}_j\\ \bar{e}_j\end{pmatrix}^T\nabla^2 f(z_j)\begin{pmatrix}-A_I^{-1}A_J\bar{e}_j\\ \bar{e}_j\end{pmatrix} \leqslant M_2(1+\|A_I^{-1}a_j\|^2)$$

令 $\xi=\min\{\|A_I^{-1}a_j\|^2, j\in J\}$,

$\eta=\max\{\|A_I^{-1}a_j\|^2, j\in J\}$

则 $$M_1(1+\xi) \leqslant \begin{pmatrix}-A_I^{-1}A_J\bar{e}_j\\ \bar{e}_j\end{pmatrix}^T\nabla^2 f(z_j)\begin{pmatrix}-A_I^{-1}A_J\bar{e}_j\\ \bar{e}_j\end{pmatrix} \leqslant M_2(1+\eta).$$

于是

$$\nabla_J f(x)^T\bar{e}_j-\nabla_I f(x)^T A_I^{-1}A_J\bar{e}_j+\frac{1}{2}hM_1(1+\xi) \leqslant \bar{s}_j \leqslant \nabla_J f(x)^T\bar{e}_j-\nabla_I f(x)^T A_I^{-1}A_J\bar{e}_j + \frac{1}{2}hM_2(1+\eta)$$

$$r(x)+\frac{1}{2}hM_1(1+\xi)\bar{e} \leqslant \bar{s} \leqslant r(x)+\frac{1}{2}hM_2(1+\eta)\bar{e}$$

即(9)成立.

由(7)，(11)有

$$\bar{r}(x)=s_J-(A_I^{-1}A_J)^T s_I = \nabla_J f(x)+\frac{1}{2}hC_J-(A_I^{-1}A_J)^T(\nabla_I f(x) + \frac{1}{2}hC_I)$$

$$\bar{r}(x)=r(x)+\frac{1}{2}hC_J-\frac{1}{2}h(A_I^{-1}A_J)^T C_I \tag{12}$$

易见 $(A_I^{-1}A_J)^T C_I$ 的分量 $(A_I^{-1}a_j)^T C_I$ 满足

$$-\sqrt{m\eta}M_2\leqslant (A_I^{-1}a_j)^T C_I\leqslant \sqrt{m\eta}M_2 \tag{12'}$$

结合(9)，(12)，(11)有

$$\bar{s}+\frac{1}{2}h[M_1-\sqrt{m\eta}M_2-M_2(1+\eta)]\bar{e} \leqslant \bar{r}(x) \leqslant \bar{s}+\frac{1}{2}h[M_2+\sqrt{m\eta}M_2-M_1(1+\xi)]\bar{e}$$

即(10)成立. 证毕

既约差商法步骤:

步 0. 选 $x^1\in R$，基 I_0，$h_0>0$，$0<\rho<1$，$k=1$，选一个十分小的终止参数 $\delta>0$.

步 1. 以 $I=I_{k-1}$ 作以下转轴运算:

1.1) 计算 $x_r^k=\min\{x_i^k : i\in I\}$，$T(I)=A_I^{-1}A$，$T_r^j$ 为 $T(I)$ 中位于 r 行和 j 列元素，$j\in J$.

1.2) 计算 $x_s^k=\max\{x_j^k : j\in J, T_r^j\neq 0\}$.

1.3) 如 $x_r^k \geqslant \rho x_\sigma^k$，令 $I_k = I, J_k = \{1,2,\cdots,n\}\backslash I_k$, $h = h_{k-1}/2$，到步 2；否则令 $I := I \cup \{\sigma\}\backslash\{r\}$，到 1.1)。

步 2. 计算差商 $\bar{s} = \bar{s}(x^k, h)$，既约差商 $\bar{r} = \bar{r}(x^k, h) = (\bar{r}_j, j \in J_k)$.

步 3. 计算搜索方向

$$d_j = \begin{cases} -\bar{r}_j, & \bar{r}j \leqslant 0 \\ -x_j^k\bar{r}_j, & \bar{r}_j > 0 \end{cases}, \quad d_{J_k} = (d_j, j \in J_k). \quad (13)$$

$$d_{I_k} = -A_{I_k}^{-1}A_{J_k}d_{J_k}, \quad d = (d_{I_K}, d_{J_K}) \quad (14)$$

步 4. 作线搜索

$$f(x^k + \bar{\lambda}d) = \min\{f(x^k + \lambda d) : x^k + \lambda d \in R\} \quad (15)$$

步 5. 如 $\bar{\lambda} = 0$，且 $h < \delta$，以 x^k 为近似解，停；如 $\bar{\lambda} = 0$，但 $h \geqslant \delta$，令 $h := \frac{1}{2}h$,，转到步 2. 如 $\bar{\lambda} \neq 0$，置 $d^k = d, \bar{r}^k = \bar{r}, \bar{s}^k = \bar{s}, h_k = h, \lambda_k = \bar{\lambda}, x^{k+1} = x^k + \lambda_k d^k$, $k := k + 1$，转回步 1.

算法的终止准则是以下面的定理 2 为理论基础.

2 算法的适定性和整体收敛性

为讨论算法的适定性和收敛性，先建立以下基本结论.

引理 1[5] 对于 x^k，步 1 中的转轴运算可有限步终止；如 $\lim\limits_{k\in K} x^k = \tilde{x}$，则 $\min\{x_j^k, k \in K, j \in I_k\} \geqslant \varepsilon_0\rho$，其中 $\varepsilon > 0$；如 $\lim\limits_{k\to\infty} x^k = \tilde{x}$，则当 k 充分大时，I_k 固定不变，即对于 $\{x^k\}$，转轴运算只需施行有限次.

引理 2 (13)，(14) 确定的方向为 (P) 在 x^k 处的可行方向；如 $r(x^k)^T d_{J_k} < 0$,，则 d 为 (P) 的下降方向，于是 $\bar{\lambda} \neq 0$.

证明 首先由假设 (a) 及[5]可知 R 的任何点 x 至少有 m 个正分量，因此当 $x_r^k \geqslant \rho\, x_\sigma^k$ 时，不难推知 $x_j^k > 0, \forall\ j \in I_k$. 另一方面，对于 $x_j^k = 0, j \in J_k$，由(13)有 $d_j \geqslant 0$. 再结合(14)可知 d 为可行方向. 由 $\nabla f(x^k)^T d = r(x^k)^T d_{J_k}$ 可知后半部分成立.

引理 3[5] x^k 为 (P) 的 $K-T$ 点，当且仅当

$$r(x^k) = r^k \geqslant 0, \quad x_{J_k}^T r^k = 0 \quad (15)$$

因 (P) 为凸规划，(15) 为 x^k 是唯一解 x^* 的充要条件.

定理 2 对于迭代点 x^k，如有正参数子列 $\{h_{kj}\}_j$，$\lim\limits_{j\to\infty} h_{kj} = 0$，使得步 5 中线搜索(15) 的步长均为零，其中 $d^{kj} = d$ 为(13)，(14) 确定的对应于参数 h_{kj} 的方向. 则 x^k 为 (P) 的最优解 x^*.

证明 因步 5 线搜索步长为零，由引理 2 有 $\nabla f(x^k)^T d^{kj} = r(x^k)^T d_{J_k}^{kj} \geqslant 0$，而

$$r(x^k)^T d_{J_k}^{kj} = \sum_{\bar{r}_i(h_{kj})\leqslant 0} -r_i(x^k)\bar{r}_i(h_{kj}) + \sum_{\bar{r}_i(h_{kj})>0} -x_i^k r_i(x^k)\bar{r}_i(h_{kj}).$$

由(12)，(12′) 有

$$\bar{r}_i(h_{kj}) - \frac{1}{2}h_{kj}M_2(1 + \sqrt{m}\eta) \leqslant r_i(x^k) \leqslant \bar{r}_i(h_{kj}) + \frac{1}{2}h_{kj}(\sqrt{m}\eta M_2 - M_1)$$

于是

$$-\sum_{\bar{r}_i(h_{kj})\leqslant 0}[\bar{r}_i(h_{kj})^2 + \frac{1}{2}h_{kj}\bar{r}_i(h_{kj})(\sqrt{m}\eta M_2 - M_1)] - \sum_{\bar{r}_i(h_{kj})>0} x_i^k\bar{r}_i(h_{kj})^2 + \sum_{\bar{r}_i(h_{kj})>0} x_i^k\bar{r}_i(h_{kj})h_{kj}M_2(1 + \sqrt{m}\eta) \geqslant r(x^k)^T d_{J_k}^{kj} \geqslant 0$$

而 $h_{kj} \to 0$，故 $\bar{r}_i(h_{kj}) \to r_i(x^k)$，于是在上式中令 $j \to \infty$ 有

$$-\sum_{r_i(x^k)\leqslant 0} r_i(x^k)^2 - \sum_{r_i(x^k)>0} x_i^k r_i(x^k)^2 \geqslant 0$$

从而 $$\sum_{r_i(x^k)\leqslant 0} r_i(x^k)^2 + \sum_{r_i(x^k)>0} x_i^k r_i(x^k)^2 = 0$$

由此可推知 $r(x^k) \geqslant 0, r(x^k)^T x_{J_k}^k = 0$. 由引理 3 之(15) 知 x^k 为 (P) 的最优解 x^*，证毕.

在下面收敛性讨论中，不妨假设步 2～步 5 对每个 x^k 可有限步终止.

定理 3 设条件 (a)，(b) 成立，则 $\{x^k\}$ 的任何极限点为 (P) 的 $K-T$ 点，从而等于最优解 x^*.

证明 不妨设 $x^k \to \tilde{x}, k \in K$，且 $I_k \equiv I$, $J_k \equiv J. \forall\ k \in K$. 利用中值定理有

$$f(x^k + h_k e_j) - f(x^k) = h_k\nabla f(x^k)^T e_j + \frac{1}{2}h_k^2 e_j^T\nabla^2 f(z^k)e_j$$

由条件 (b) 知

$$h_k\nabla f(x^k)^T e_j + \frac{1}{2}h_k^2 M_1 \leqslant f(x^k + h_k e_j) - f(x^k) \leqslant h_k\nabla f(x^k)^T e_j + \frac{1}{2}h_k^2 M_2.$$

故

$$\nabla f(x^k)^T e_j + \frac{1}{2}h_k M_1 \leqslant s_j^k = \frac{f(x^k + h_k e_j) - f(x^k)}{h_k} \leqslant \nabla f(x^k)^T e_j + \frac{1}{2}h_k M_2.$$

由 $h_k \to 0, (k \to \infty)$，有

$$\lim_{K\in K} s_j^k = \nabla f(\tilde{x})^T e_j = \frac{\partial f(\tilde{x})}{\partial x_j}, j = 1 \sim n.$$

即 $\lim\limits_{k\in K}s^k = \nabla f(\tilde{x})$. 因此 $\lim\limits_{k\in K}\bar{r}^k = \lim\limits_{k\in K}\bar{r}(x^k,h_k)$

$= \lim\limits_{k\in K}[s_J^k - (A_I^{-1}A_J)^T s_I^k]$

$= \nabla_J f(\tilde{x}) - (A_I^{-1}A_J)^{-1}\nabla_I f(\tilde{x}) = r(\tilde{x})$.

引进方向 $\tilde{d}$:

$$\tilde{d}_j = \begin{cases} -r_j(\tilde{x}), r_j(\tilde{x}) \leqslant 0 \\ -\tilde{x}_j r_j(\tilde{x}), r_j(\tilde{x}) > 0 \end{cases}, j \in J.$$

$\tilde{d}_I = -A_I^{-1}A_J\tilde{d}_J$.

则 $d^k \to \tilde{d}, k \in K$. 下面假设 $\tilde{x}$ 不是 $K-T$ 点，我们将导出矛盾.

首先由引理1,3有

$$\nabla f(\tilde{x})^T\tilde{d} = r(\tilde{x})^T\tilde{d}_J = -\sum_{r_j(\tilde{x})\leqslant 0} r_j(\tilde{x})^2 - \sum_{r_j(\tilde{x})>0}\tilde{x}_j r_j(\tilde{x})^2 < 0.$$

由 $x^k \to \tilde{x}, d^k \to \tilde{d}, k \in K$ 以及 $\nabla f(x)$ 的一致连续性可推知:存在 $\varepsilon' > 0$,使得 $k \in K$ 充分大时,

$$\nabla f(x^k + \lambda d^k)^T d^k < -\delta, \forall \lambda \in [0,\varepsilon'], \quad (16)$$

其中 $\delta = -\frac{1}{2}\nabla f(\tilde{x})^T\tilde{d} > 0$.

记

$$\varepsilon_k = \begin{cases} \min\left\{\frac{-x_j^k}{d_j^k}: d_j^k < 0\right\} \\ +\infty, \text{如 } d_j^k \geqslant 0, j = 1 \sim n. \end{cases}$$

则 $\qquad x^k + \lambda d^k \in R, \lambda \in [0,\varepsilon_k]$.

其次我们证明:$\varepsilon = \inf\{\varepsilon_k: k \in K\} > 0$.

若不然,可取子列 $K' \subset K$,使得

$\varepsilon_k = \frac{-x_t^k}{d_t^k} \to 0, d_t^k < 0, k \in K', t$ 与 k 无关. 由于 $x_t^k \to \tilde{x}_t, d_t^k \to \tilde{d}_t$,故 $\tilde{x}_t = 0, t \in J$,(引理1). 结合(14)及 $d_t^k < 0$ 可知 $d_t^k = -x_t^k\bar{r}_t^k$. 于是

$$\varepsilon_k = \frac{-x_t^k}{d_t^k} = \frac{1}{\bar{r}_t^k}.$$

但 $\bar{r}_t^k \to r_t(\tilde{x})$,这与 $\varepsilon_k \to 0$ 矛盾. 即 $\varepsilon > 0$.

由于 $\varepsilon_k \geqslant \varepsilon > 0$,故

$$x^k + \lambda d^k \in R, \forall \lambda \in [0,\varepsilon], k \in K \quad (17)$$

最后我们来导出矛盾.

令 $\bar{\varepsilon} = \min\{\varepsilon,\varepsilon'\} > 0$. 由中值定理及(15),(17)

$$f(x^{k+1}) = f(x^k + \lambda_k d^k) \leqslant f(x^k + \bar{\varepsilon}d^k) = f(x^k) + \nabla f(x^k + \theta_k\bar{\varepsilon}d^k)^T(\bar{\varepsilon}d^k). \quad \theta_k \in [0,1],$$

故 $\theta_k\bar{\varepsilon} \in [0,\varepsilon']$,由(16)有

$$f(x^{k+1}) \leqslant f(x^k) - \bar{\varepsilon}\delta, k \in K.$$

由 $\{f(x^k)\}$ 单调不增及 $f(x^k) \to f(\tilde{x}), k \in K$ 有 $f(x^k) \to f(\tilde{x}) (k \to \infty)$. 在上不等式中令 $k \to \infty, k \in K$ 有 $\bar{\varepsilon}\delta \leqslant 0$. 这矛盾.

因此 $\tilde{x}$ 为(P)的 $K-T$ 点,从而 $\tilde{x}$ 等于凸规划(P)的唯一最优解 x^*,证毕.

推论 如 $\{x^k\}$ 有界,则 $x^k \to x^*, k \to \infty$.

算法的几点说明:

1) 算法还可以采用别的终止准则,如"$\bar{r}(x^k,h)^T d_{J_k} \geqslant 0, h < \delta$". 至于更为优秀的终止准则有待进一步研究和探索.

2) 为减少计算量,可以考虑将精确线搜索(15)改为以下非精确线搜索. 收敛性仍成立.

取 $\bar{\lambda}$ 为 $\{1,\frac{1}{2},\frac{1}{4},\frac{1}{8},\cdots,\frac{1}{2^j},\cdots,0\}$ 中满足式(18)的最大者

$$\begin{cases} f(x^k + \lambda d) \leqslant f(x^k) + \frac{1}{2}\lambda\bar{r}(x^k,h)^T d_{J_k} \\ f(x^k + \lambda d) \leqslant f(x^k) \end{cases} \quad (18)$$

如式(18)对非常大的 $j, \lambda = 2^{-j}$ 仍不成立,可取步长 $\bar{\lambda} = 0$. (15)中不要求 $\lambda \geqslant 0$ 是为了尽快使 $\bar{\lambda} \neq 0$. 结束一次迭代. 当有某个 $x_j^k = 0$,且 $\bar{r}_j < 0$ 时,为使 $x^k + \lambda d \in R$,自然有 $\lambda \geqslant 0$.

3) 利用本文思想,类似于文献[5]可建立包含许多算法在内的一类既约差商法.

参考文献

1 Wolfe P. An Extended Simplex Method. Notices of the Anerican Mathematical Society, 1962, 9: 4.

2 Laslon L S. Reduced Gradient Methods, Prented at NATO Advanced Research Institute on Non-Linear Optimization. Cambridge, England, July 1981, 13~14.

3 Dixon L C. On the Convergence of Variable Metric Method with Numerical Derivative and Effect of Noise in the Function Evaluation. In: Oettli V and Ritler K ed., Optimization and Operations Research, Springer-Verlong, 1976.

4 赵小平. 差商变尺度法的整体收敛性. 应用数学, 1994, 7 (1): 41~47.

5 越民义, 韩继业, 姚恩瑜. 可行方向法的一个统一探讨. 数学年刊, 1985, 6A (1): 1~12.

(责任编辑: 莫鼎新)