

## 非单调广义对角拟牛顿算法

周群艳<sup>1</sup>, 杭丹<sup>2</sup>

(1. 江苏理工学院数理学院, 江苏 常州 213001)

(2. 空军勤务学院基础部, 江苏 徐州 221000)

**摘要:** 本文研究了无约束最优化的求解问题. 利用新的对角拟牛顿校正和非单调技术, 获得了一种非单调广义对角拟牛顿算法. 新算法具有低存储、低计算量的特点, 非常适合大规模问题的求解, 推广了文献 [8] 的结果.

**关键词:** 弱拟牛顿方程; 对角校正; 非单调技术; 全局收敛性; 数值实验

MR(2010) 主题分类号: 65K05; 90C30      中图分类号: O224

文献标识码: A      文章编号: 0255-7797(2016)02-0335-11

### 1 引言

考虑无约束最优化问题

$$\min_{x \in R^n} f(x), \quad (1.1)$$

其中  $f(x) : R^n \rightarrow R$  为连续可微函数.

拟牛顿法是解 (1.1) 的最常用的方法之一, 它以拟牛顿方程为基础, 利用目标函数一阶导数的信息, 构造出目标函数的曲率近似, 从而避免了求 Hesse 阵或其逆矩阵. 记  $g_k = \nabla f(x_k)$ , 其基本迭代格式为

$$x_{k+1} = x_k + \alpha_k d_k, \quad (1.2)$$

其中  $\alpha_k$  为通过某种搜索策略确定的步长因子,  $d_k$  为搜索方向  $-H_k g_k$ ,  $H_k$  为目标函数  $f(x)$  的 Hesse 阵的逆的近似, 要求正定且满足拟牛顿方程

$$H_{k+1} y_k = s_k, \quad (1.3)$$

其中  $y_k = g_{k+1} - g_k$ ,  $s_k = x_{k+1} - x_k$ .

传统的确定步长因子  $\alpha_k$  的 Armijo 线搜索策略要求  $\alpha_k$  满足

$$f(x_k + \alpha_k d_k) \leq f(x_k) + \gamma \alpha_k g_k^T d_k, \quad (1.4)$$

其中  $\gamma \in (0, 1)$ . 显然这样搜索要求每次迭代目标函数值单调下降, 大大降低了算法的效率<sup>[1]</sup>. 非单调线搜索技术由于其有利于求解全局最优解和算法的快速收敛而受到许多最优化爱好

\*收稿日期: 2013-07-21      接收日期: 2014-05-16

基金项目: 江苏省高校自然科学研究项目 (13KJB110007); 江苏理工学院基础及应用基础研究项目 (KYY13012); 江苏理工学院博士启动基金项目 (KYY13005).

作者简介: 周群艳 (1980-), 女, 江苏常州, 讲师, 主要研究方向: 最优化理论与方法.

者的关注. Grippo 等在文献 [1] 中提出的非单调线搜索技术放宽了对  $\alpha_k$  的选取范围, 每次迭代仅要求步长因子  $\alpha_k$  满足下式

$$f(x_k + \alpha_k d_k) \leq f(x_{l(k)}) + \gamma \alpha_k g_k^T d_k, \quad (1.5)$$

其中  $f(x_{l(k)}) = \max_{0 \leq j \leq m_k} \{f(x_{k-j})\}$ ,  $m_0 = 0$ ,  $0 \leq m_k \leq \min\{m_{k-1} + 1, M\}$  ( $k \geq 1$ ),  $M \geq 0$  为正整数. 这种非单调技术在文献 [2] 中被推广到信赖域算法. 但是 Grippo 等人的非单调技术有一些不足之处, 比如从式 (1.5) 可见, 由于  $f(x_{l(k)})$  是前面  $M$  个函数值里的最大者, 所以可能导致某步迭代得到的较好的函数值信息没有被用上, 另外算法的数值表现受  $M$  的取值的影响. 于是, Zhang 等在文献 [3] 中提出了一种新的非单调线搜索方法, 将式 (1.5) 中的  $f(x_{l(k)})$  换成某些函数值的加权平均  $C_k$ , 即步长因子  $\alpha_k$  要满足

$$f(x_k + \alpha_k d_k) \leq C_k + \gamma \alpha_k g_k^T d_k, \quad (1.6)$$

其中

$$C_k = \begin{cases} f(x_k), & k = 0, \\ \frac{\eta_{k-1} Q_{k-1} C_{k-1} + f(x_k)}{Q_k}, & k \geq 1, \end{cases} \quad Q_k = \begin{cases} 1, & k = 0, \\ \eta_{k-1} Q_{k-1} + 1, & k \geq 1, \end{cases}$$

$\eta_{k-1} \in [\eta_{\min}, \eta_{\max}]$ ,  $\eta_{\min} \in [0, 1)$  及  $\eta_{\max} \in [\eta_{\min}, 1)$  为两个选定的参数. 文献 [4] 中提出了采用该非单调策略的信赖域算法. 随后, 文献 [5] 中又指出上述非单调方法每次迭代要计算  $Q_k$  以及  $C_k$ , 这在计算上是一种浪费, 于是提出了一种新的非单调信赖域方法, 将 (1.6) 式中的  $C_k$  替换为  $D_k$ , 这里  $D_k$  是  $D_{k-1}$  和  $f_k$  的一个简单凸组合, 即

$$D_k = \begin{cases} f_k, & k = 1, \\ \eta_k D_{k-1} + (1 - \eta_k) f_k, & k \geq 2, \end{cases} \quad (1.7)$$

其中  $\eta_k \in (0, 1)$  可以是变量, 也可以取常数. 数值试验表明文献 [5] 中的非单调策略更高效.

通常拟牛顿法每次迭代需要存储一个矩阵  $H_k$ , 而该矩阵一般是稠密的, 即算法的存储量至少为  $O(n^2)$ , 对中小规模的优化问题有较好的数值表现. 一般来说, 大规模优化问题的目标函数 Hesse 阵是对称的, 且具有某种稀疏结构. 经典拟牛顿法 (如 DFP 方法、BFGS 方法), 虽然校正矩阵能继承正定性和对称性, 却不能继承其稀疏性, 存储空间问题成为阻碍拟牛顿方法求解大规模问题的一大障碍. 因此, 希望得到存储量小且收敛速度快、计算量小的算法, 需对传统的拟牛顿法进行改进. 占据主要内存单元的 Hesse 阵是重点考虑对象, 若限制拟牛顿法中 Hesse 阵或其逆的近似为对角矩阵则可以大大降低存储量, 在一定的条件下也能保持相对较快的收敛速度.

1988 年, Barzilai 和 Borwein<sup>[6]</sup> 提出的两点步长梯度法 (BB 方法), 利用当前迭代点以及前一点的信息来确定梯度方向的步长, 也可看成用具有一定拟牛顿性质的数量矩阵  $H_{k+1} = \alpha_{k+1} I$  作为 Hesse 阵的逆的近似, 其中  $\alpha_{k+1}$  为  $\alpha_{k+1} = \operatorname{argmin}_{\alpha} \|\alpha y_k - s_k\|^2$ , 即

$$\alpha_{k+1} = \frac{s_k^T y_k}{y_k^T y_k}.$$

BB 方法对二维的二次严格凸的目标函数是超线性收敛的, 对于目标函数是一般的二次凸函数则是 R - 线性收敛的, 但不能保证全局收敛性. 为克服这个缺点, 同时保留 BB 步长的

性质, Raydan<sup>[7]</sup> 将 BB 方法与 Grippo 等人的非单调线搜索方法<sup>[1]</sup> 结合, 构造了一个具有全局收敛性的非单调线搜索算法 (GBB 方法). 数值试验结果表明 GBB 方法对于求解大规模无约束优化问题的数值表现较好. 2006 年时贞军<sup>[8]</sup> 受 BB 方法的启发, 在拟牛顿法中, 限制 Hesse 阵逆阵的近似  $H_{k+1}$  为正定对角矩阵且近似满足经典的拟牛顿方程 (1.3), 通过求解

$$\min_{\underline{h} \leq h_{k+1}^{(i)} \leq \bar{h}} \|H_{k+1}y_k - s_k\|^2$$

而得到  $H_{k+1}$  (其中  $\underline{h}$  和  $\bar{h}$  为两个正数), 即  $H_{k+1}$  的第  $i$  个对角元

$$h_{k+1}^{(i)} = \begin{cases} \frac{s_k^{(i)}}{y_k^{(i)}}, & \text{若 } \underline{h} \leq \frac{s_k^{(i)}}{y_k^{(i)}} \leq \bar{h}; \\ \underline{h}, & \text{若 } \frac{s_k^{(i)}}{y_k^{(i)}} < \underline{h}; \\ \bar{h}, & \text{若 } \frac{s_k^{(i)}}{y_k^{(i)}} > \bar{h}; \\ h_k^i, & \text{若 } y_k^{(i)} = 0. \end{cases} \quad (1.8)$$

从而构造了对角稀疏拟牛顿法, 该方法为求解大规模问题提供了新的思路. 文献 [9, 10] 在此基础上设计了求解无约束最优化问题的非单调对角稀疏拟牛顿算法.

以上的对角拟牛顿法都是建立在经典拟牛顿方程 (1.3) 的基础之上. 近年来, 基于修正拟牛顿方程的修正拟牛顿法的研究吸引了不少国内外学者. 1991 年, Yuan<sup>[11]</sup> 提出了一种修正 BFGS 方法, 随后, Zhang 和 Xu<sup>[12]</sup>, Wei、Li 和 Qi<sup>[13]</sup> 等相继又提出了几类修正拟牛顿算法. 理论分析和数值试验表明, 这些改进后的拟牛顿法由于利用了迭代点的更多信息, 所以比经典拟牛顿法的数值表现更好. 为获取 Hesse 阵的逆矩阵的更良好的对角逼近, 本文以逼近程度更好的修正拟牛顿方程为基础, 推导出一类新的对角校正公式, 并将文献 [5] 中的非单调策略应用于线搜索算法框架, 设计了一种求解无约束优化问题的广义对角拟牛顿算法.

## 2 新的对角校正及广义对角拟牛顿算法

考虑方程

$$B_{k+1}s_k = t_k y_k \quad (t_k > 0), \quad (2.1)$$

其中  $B_{k+1}$  为 Hesse 阵的近似矩阵. 当  $t_k = 1$  时, 式 (2.1) 为经典拟牛顿方程. 当  $t_k$  选取适当的值时, 可以得到相关的修正拟牛顿方程<sup>[11]</sup>, 比如可取

$$t_k = t_k^* = \frac{2}{s_k^T y_k} (f_k - f_{k+1} + s_k^T g_{k+1})$$

或

$$t_k = t_k^{**} = \frac{1}{s_k^T y_k} [(g_{k+1} - g_k)^T s_k + 6(f_k - f_{k+1}) + 3(g_k + g_{k+1})^T s_k].$$

易证得当  $t_k = t_k^*$  或  $t_k = t_k^{**}$  时, 以式 (2.1) 为基础得到的  $B_{k+1}$  较  $t_k = 1$  时得到的  $B_{k+1}$  在某种意义上更逼近精确的 Hesse 阵 (见文献 [12, 13]).

由式 (2.1) 得, Hesse 阵的近似逆阵  $H_{k+1}$  需满足如下修正拟牛顿方程

$$H_{k+1}y_k = \frac{1}{t_k} s_k. \quad (2.2)$$

本文中要求正定对角阵近似满足相应的弱拟牛顿方程<sup>[14]</sup>

$$y_k^T H_{k+1} y_k = \frac{1}{t_k} y_k^T s_k \equiv \rho_k. \quad (2.3)$$

事实上, 满足式 (2.3) 的对角矩阵一般有无数多个, 在其中选取一个最近似满足经典拟牛顿方程 (1.3) 的一个, 即有如下子问题

$$\begin{aligned} & \min \frac{1}{2} \|H_{k+1} y_k - s_k\|^2, \\ & \text{s.t. } y_k^T H_{k+1} y_k = \rho_k, \\ & H_{k+1} \text{ 为对角矩阵,} \end{aligned} \quad (2.4)$$

其中  $\|\cdot\|$  为欧氏范数. 然而, 问题 (2.4) 的解未必正定, 为保证其正定性, 需进一步限制  $h_{k+1}^{(i)}$  在一定的范围内, 即要求  $0 < \underline{h} \leq h_{k+1}^{(i)} \leq \bar{h}$ . 下面给出新的正定对角阵的具体构造方法.

问题 (2.4) 的目标函数是凸函数, 且约束集为凸集, 其 Lagrangian 函数为

$$L(H_{k+1}, \mu) = \frac{1}{2} \|H_{k+1} y_k - s_k\|^2 + \mu(\rho_k - y_k^T H_{k+1} y_k), \quad (2.5)$$

其中  $\mu$  为对应于约束条件的 Lagrangian 乘子. 将  $L$  关于  $H_{k+1}$  的每个对角元求偏导数, 并置其为 0, 得

$$h_{k+1}^{(i)} (y_k^{(i)})^2 = \mu (y_k^{(i)})^2 + s_k^{(i)} y_k^{(i)} \quad (i = 1, 2, \dots, n). \quad (2.6)$$

将式 (2.6) 中各式相加, 结合约束条件  $y_k^T H_{k+1} y_k = \rho_k$ , 得

$$\mu = \frac{\rho_k - s_k^T y_k}{y_k^T y_k}. \quad (2.7)$$

将式 (2.7) 代入式 (2.6) 得, 当  $y_k^{(i)} \neq 0$  时,

$$h_{k+1}^{(i)} = \frac{\rho_k - s_k^T y_k}{y_k^T y_k} + \frac{s_k^{(i)}}{y_k^{(i)}}. \quad (2.8)$$

若记  $\bar{h}_{k+1}^{(i)} = \frac{\rho_k - s_k^T y_k}{y_k^T y_k} + \frac{s_k^{(i)}}{y_k^{(i)}}$ , 则

$$h_{k+1}^{(i)} = \begin{cases} \bar{h}_{k+1}^{(i)}, & \text{若 } \underline{h} \leq \bar{h}_{k+1}^{(i)} \leq \bar{h}; \\ \underline{h}, & \text{若 } \bar{h}_{k+1}^{(i)} < \underline{h}; \\ \bar{h}, & \text{若 } \bar{h}_{k+1}^{(i)} > \bar{h}; \\ h_k^i, & \text{若 } y_k^{(i)} = 0. \end{cases} \quad (2.9)$$

这样便得到了近似满足弱拟牛顿方程 (2.3) 的正定对角矩阵  $H_{k+1}$ . 显然, 若  $t_k = 1$ , 即  $\rho_k = s_k^T y_k$  时, 校正公式 (2.9) 退化为文献 [8] 中的公式 (1.8).

接下来给出求解无约束优化问题的非单调广义对角拟牛顿算法.

#### 非单调广义对角拟牛顿算法 (GDQN)

**步 0** 给出初始点  $x_0 \in R^n$ ,  $0 < \gamma < 1$ ,  $0 < \beta < 1$ ,  $0 < \underline{h} < \bar{h}$ ,  $0 \leq \eta_{\min} < \eta_{\max} < 1$ . 置  $k = 0$ ,  $H_0 = I$ ,  $D_0 = f(x_0)$ .

步 1 若满足算法终止准则, 则停止, 否则, 转步 2.

步 2 计算  $d_k = -H_k g_k$ .

步 3 选取  $\alpha_k$  为  $\{1, \beta, \beta^2, \dots\}$  中满足下式的最大者

$$f(x_k + \alpha d_k) \leq D_k + \gamma \alpha g_k^T d_k, \quad (2.10)$$

令  $x_{k+1} = x_k + \alpha_k d_k$ .

步 4 选取  $\eta_{k+1} \in [\eta_{\min}, \eta_{\max}]$ , 按 (1.7) 式计算  $D_{k+1}$ .

步 5 计算  $s_k, y_k$  及  $\rho_k$ , 按 (2.9) 式校正  $H_{k+1}$ , 令  $k = k + 1$ , 转步 1.

注 (1) 若  $\eta_k$  恒取 0 且  $\rho_k = s_k^T y_k$ , 算法 GDQN 即为文献 [8] 中的单调对角稀疏拟牛顿算法.

(2) 按照文献 [11] 中的修正拟牛顿方程的构造方法,  $\rho_k$  的选取可按如下两种方式选取:

(a)  $t_k = \frac{2}{s_k^T y_k} (f_k - f_{k+1} + s_k^T g_{k+1})$ , 即

$$\rho_k = \frac{(s_k^T y_k)^2}{2(f_k - f_{k+1} + s_k^T g_{k+1})}. \quad (2.11)$$

(b)  $t_k = \frac{1}{s_k^T y_k} [(g_{k+1} - g_k)^T s_k + 6(f_k - f_{k+1}) + 3(g_k + g_{k+1})^T s_k]$ , 即

$$\rho_k = \frac{(s_k^T y_k)^2}{(g_{k+1} - g_k)^T s_k + 6(f_k - f_{k+1}) + 3(g_k + g_{k+1})^T s_k}. \quad (2.12)$$

对一般的函数, 由式 (2.11) 或 (2.12) 定义的  $\rho_k$  可能是负的, 甚至可能溢出, 为解决这一问题, 按如下安全方式截取  $\rho_k$ ,

$$\rho_k = \begin{cases} \underline{h} \|y_k\|^2, & \text{如果 } \rho_k < \underline{h} \|y_k\|^2; \\ \bar{h} \|y_k\|^2, & \text{如果 } \rho_k > \bar{h} \|y_k\|^2. \end{cases}$$

### 3 全局收敛性分析

本节分析算法 GDQN 的收敛性, 首先作出如下假设:

假设 3.1 (1) 水平集  $L(x_0) = \{x \mid f(x) \leq f(x_0)\}$  有界; (2)  $f(x)$  在  $L(x_0)$  上二次连续可微.

引理 3.2 若  $x_k$  不是函数  $f(x)$  的稳定点, 则 (1)  $\|d_k\| \leq \bar{h} \|g_k\|$ ; (2)  $g_k^T d_k \leq -\underline{h} \|g_k\|^2$ .

证 由  $d_k$  的定义和  $H_k$  的取法易知.

引理 3.3 假设  $\{x_k\}$  是算法 GDQN 产生的序列, 则对任何  $k$ , 有

$$D_{k+1} \geq f_{k+1}. \quad (3.1)$$

证 由  $D_{k+1}$  的定义得

$$D_{k+1} - f_{k+1} = \eta_{k+1} (D_k - f_{k+1}). \quad (3.2)$$

由引理 3.2 及线搜索准则 (2.10) 易得  $D_k - f_{k+1} \geq -\gamma \alpha_k g_k^T d_k \geq 0$ . 因此由式 (3.2) 即得式 (3.1) 成立.

**引理 3.4** 算法 GDQN 中步 3 中的线搜索必在有限步内终止, 即算法 GDQN 是适定的.

**证** 用反证法. 假设结论不成立, 即存在某个  $k$ , 使得对于任何正整数  $i$ , 有

$$f(x_k + \beta^i d_k) > D_k + \gamma \beta^i g_k^T d_k. \quad (3.3)$$

由引理 3.3 得  $f_k \leq D_k$ . 于是结合式 (3.3) 得

$$\frac{f(x_k + \beta^i d_k) - f_k}{\beta^i} > \gamma g_k^T d_k. \quad (3.4)$$

由于  $f(x)$  二次连续可微, 令  $i \rightarrow \infty$ , 将式 (3.4) 两边取极限得  $g_k^T d_k \geq \gamma g_k^T d_k$ . 该式表明  $g_k^T d_k \geq 0$ , 与引理 3.2 矛盾.

**引理 3.5** 假设  $\{x_k\}$  是算法 GDQN 产生的序列, 则对任何  $k$ , 存在  $M > 0$ , 使得步长因子  $\alpha_k$  满足

$$\alpha_k > \min\left\{\frac{\beta}{2}, \frac{2\beta(\gamma-1)g_k^T d_k}{M\|d_k\|^2}\right\}. \quad (3.5)$$

**证** 若  $\alpha_k/\beta > \frac{1}{2}$ , 则  $\alpha_k > \frac{\beta}{2}$ . 以下考虑  $\alpha_k/\beta \leq \frac{1}{2}$  的情形. 由算法 GDQN 的步 3 得

$$f(x_k + \frac{\alpha_k}{\beta} d_k) > D_k + \gamma \frac{\alpha_k}{\beta} g_k^T d_k. \quad (3.6)$$

由 Taylor 展开式得

$$f(x_k + \frac{\alpha_k}{\beta} d_k) = f_k + \frac{\alpha_k}{\beta} g_k^T d_k + \frac{1}{2} \left(\frac{\alpha_k}{\beta}\right)^2 d_k^T \nabla^2 f(\xi_k) d_k, \quad (3.7)$$

其中  $\xi_k \in (x_k, x_k + \frac{\alpha_k}{\beta} d_k)$ .

由于  $f(x)$  在有界闭集  $L(x_0)$  上二次连续可微, 所以存在常数  $M > 0$ , 使得  $\|\nabla^2 f(x)\| \leq M$ . 结合式 (3.1), (3.6), (3.7) 以及假设 3.1, 可得

$$f_k + \gamma \frac{\alpha_k}{\beta} g_k^T d_k < f_k + \frac{\alpha_k}{\beta} g_k^T d_k + \frac{1}{2} \left(\frac{\alpha_k}{\beta}\right)^2 d_k^T \nabla^2 f(\xi_k) d_k \leq f_k + \frac{\alpha_k}{\beta} g_k^T d_k + \frac{1}{2} \left(\frac{\alpha_k}{\beta}\right)^2 M \|d_k\|^2.$$

因此

$$\alpha_k > \frac{2\beta(\gamma-1)g_k^T d_k}{M\|d_k\|^2}.$$

于是式 (3.5) 得证.

**引理 3.6** 假设  $\{x_k\}$  是算法 GDQN 产生的序列, 则序列  $\{D_k\}$  单调不增.

**证** 由线搜索准则 (2.10)、引理 3.2 和引理 3.5 得

$$\begin{aligned} f_{k+1} &\leq D_k + \gamma \alpha_k g_k^T d_k \leq D_k - \gamma h \min\left\{\frac{\beta}{2}, \frac{2\beta(\gamma-1)g_k^T d_k}{M\|d_k\|^2}\right\} \|g_k\|^2 \\ &\leq D_k - \gamma \min\left\{\frac{\beta h}{2}, \frac{2\beta(1-\gamma)h^2}{Mh^2}\right\} \|g_k\|^2 \equiv D_k - \delta \|g_k\|^2, \end{aligned} \quad (3.8)$$

其中常数  $\delta = \gamma \min\left\{\frac{\beta h}{2}, \frac{2\beta(1-\gamma)h^2}{Mh^2}\right\}$ . 根据  $D_{k+1}$  的定义及式 (3.8) 有

$$\begin{aligned} D_{k+1} &= \eta_{k+1} D_k + (1 - \eta_{k+1}) f_{k+1} \\ &\leq \eta_{k+1} D_k + (1 - \eta_{k+1}) (D_k - \delta \|g_k\|^2) \\ &= D_k - (1 - \eta_{k+1}) \delta \|g_k\|^2. \end{aligned} \quad (3.9)$$

(3.9) 式表明序列  $\{D_k\}$  单调不增.

**引理 3.7** 若假设 3.1 成立,  $\{x_k\}$  是算法 GDQN 产生的序列, 则序列  $\{D_k\}$  收敛.

**证** 由假设 3.1, 引理 3.3, 引理 3.6 及  $D_0 = f(x_0)$  易得算法 GDQN 产生的序列  $\{x_k\}$  包含在有界水平集  $L(x_0)$  中, 于是进一步可得  $\{D_k\}$  收敛.

**定理 3.8** 若假设 3.1 成立, 算法 GDQN 产生的无穷点列为  $\{x_k\}$ , 则

$$\liminf_{k \rightarrow \infty} \|g_k\| = 0. \quad (3.10)$$

**证** 用反证法. 假设式 (3.10) 不成立, 则存在  $\varepsilon > 0$  使得对任何  $k$ , 有

$$\|g_k\| \geq \varepsilon. \quad (3.11)$$

结合式 (3.9) 和 (3.11) 得

$$D_k - D_{k+1} \geq (1 - \eta_{k+1})\delta\|g_k\|^2 \geq (1 - \eta_{k+1})\delta\varepsilon^2. \quad (3.12)$$

于是

$$\sum_{k=1}^{\infty} (D_k - D_{k+1}) \geq \sum_{k=1}^{\infty} (1 - \eta_{k+1})\delta\varepsilon^2. \quad (3.13)$$

由于  $\{D_k\}$  收敛, 因此

$$+\infty > \sum_{k=1}^{\infty} (1 - \eta_{k+1})\delta\varepsilon^2 \geq \sum_{k=1}^{\infty} (1 - \eta_{\max})\delta\varepsilon^2. \quad (3.14)$$

而事实上  $\sum_{k=1}^{\infty} (1 - \eta_{\max})\delta\varepsilon^2$  发散, 这与 (3.14) 式矛盾, 因此该定理成立.

## 4 数值实验

为验证算法 GDQN 的有效性, 本节选用文献 [15] 中的一些测试函数, 利用 MATLAB 编制程序进行数值实验, 并与采用文献 [8] 中的校正方法对应的算法进行比较.

算法 GDQN 中的有关参数选取如下:  $\gamma = 0.0001$ ,  $\beta = 0.5$ ,  $\eta_k$  取常数 0.5, 选取

$$\begin{aligned} \underline{h} &= \max\left\{0.5 \cdot \frac{|(x_k - x_{k-1})^T(g_k - g_{k-1})|}{\|g_k - g_{k-1}\|^2}, 0.0001\right\}, \\ \bar{h} &= \min\left\{5 \cdot \frac{|(x_k - x_{k-1})^T(g_k - g_{k-1})|}{\|g_k - g_{k-1}\|^2}, 10000\right\}. \end{aligned}$$

根据算法 GDQN 中的  $\rho_k$  的不同取法, 得到如下的三个不同的算法:

- (1) DQN:  $\rho_k = s_k^T y_k$ , 即采用文献 [8] 中的校正方法的非单调对角拟牛顿法.
- (2) GDQN1:  $\rho_k$  由式 (2.11) 确定.
- (3) GDQN2:  $\rho_k$  由式 (2.12) 确定.

算法的终止准则为  $\|g_k\|_{\infty} \leq 10^{-5}(1 + |f(x_k)|)$ .

当迭代次数超过 5000 时迭代也终止. 计算结果见表 1. 表中 NF 表示函数估值次数, Iter 表示迭代次数, CPU 表示计算机执行算法所需的时间.

表 1: 算法 DQN, GDQN1 和 GDQN2 的数值结果

函数名称	$n$	DQN			GDQN1			GDQN2		
		NF	Iter	CPU	NF	Iter	CPU	NF	Iter	CPU
Perturbed Quadratic	100	25	18	0.0129	25	18	0.0157	25	18	0.0150
	1000	43	37	0.0537	43	37	0.0625	43	37	0.0583
	5000	90	71	0.4133	90	71	0.3999	89	71	0.4132
Almost Perturbed Quadratic	100	15	8	0.0127	15	8	0.0153	15	8	0.0150
	1000	20	10	0.0161	20	10	0.0191	20	10	0.0207
	5000	30	18	0.1390	31	19	0.1508	32	20	0.1494
Extended Powell	100	506	477	0.2770	419	395	0.2454	299	273	0.1728
	1000	919	889	1.6545	692	667	1.3822	900	862	1.6519
	5000	1056	1017	6.7567	662	633	4.6871	857	809	5.6442
	10000	736	692	8.7139	743	713	9.3528	891	848	11.0917
Extended Rosenbrock	100	165	148	0.0762	123	113	0.0525	97	90	0.0442
	1000	166	148	0.1266	123	113	0.0987	118	114	0.1181
	5000	135	126	0.4596	75	72	0.3010	87	81	0.3426
	10000	193	175	1.0611	174	165	1.0473	111	105	0.6596
Raydan1	100	7	7	0.0080	6	6	0.0062	6	5	0.0056
	1000	7	7	0.0197	6	6	0.0156	6	5	0.0110
	5000	7	7	0.0422	6	6	0.0482	6	5	0.0391
	10000	7	7	0.0757	6	6	0.0701	6	5	0.0679
Raydan2	100	15	12	0.0075	14	11	0.0085	14	10	0.0078
	1000	17	11	0.0232	17	11	0.0210	17	10	0.0200
	5000	19	10	0.0880	17	9	0.0638	18	9	0.0690
	10000	20	10	0.1639	18	9	0.1119	19	9	0.1254
Broyden Tridiagonal	100	139	111	0.0806	95	74	0.0585	97	76	0.0596
	1000	119	93	0.2278	88	67	0.1988	97	76	0.2085
	5000	104	80	0.8401	88	67	0.7287	126	104	1.2740
	10000	104	80	0.8502	88	67	0.7121	112	90	1.0642
Diagonal 1	100	19	15	0.0170	17	13	0.0108	17	13	0.0104
	1000	23	14	0.0274	21	12	0.0227	22	13	0.0242
	5000	24	13	0.1042	23	12	0.0937	23	12	0.0748
	10000	44	26	0.3568	41	24	0.3295	51	25	0.3546
Diagonal 2	100	18	18	0.0131	15	15	0.0129	17	16	0.0133
	1000	34	31	0.0670	28	27	0.0512	33	28	0.0564
	5000	58	48	0.3073	36	34	0.2018	42	36	0.2374
	10000	76	51	0.6480	63	49	0.6123	61	45	0.5922
Diagonal 3	100	15	9	0.0083	15	9	0.0092	15	9	0.0091
	1000	17	8	0.0286	17	8	0.0304	15	6	0.0189
	5000	19	8	0.0763	19	8	0.0799	17	6	0.0733
	10000	20	8	0.1396	20	8	0.1859	18	6	0.1017
Diagonal 4	100	11	5	0.0029	11	5	0.0054	11	5	0.0046
	1000	11	5	0.0104	11	5	0.0137	11	5	0.0140
	5000	11	5	0.0244	11	5	0.0285	11	5	0.0298
	10000	11	5	0.0466	11	5	0.0551	11	5	0.0505
Diagonal 5	100	5	5	0.0040	5	5	0.0048	5	5	0.0048
	1000	5	5	0.0141	5	5	0.0140	4	4	0.0113
	5000	5	5	0.0389	5	5	0.0447	4	4	0.0339
	10000	5	5	0.0707	4	4	0.0591	4	4	0.0584

表 1(续): 算法 DQN, GDQN1 和 GDQN2 的数值结果

函数名称	$n$	DQN			GDQN1			GDQN2		
		NF	Iter	CPU	NF	Iter	CPU	NF	Iter	CPU
Dixon	100	108	100	0.0627	128	122	0.0728	96	88	0.0543
	1000	103	98	0.1875	139	131	0.2563	105	95	0.1593
	5000	118	112	0.6787	149	140	0.8560	107	98	0.6000
	10000	122	115	1.3880	142	138	1.8088	108	100	1.2561
Hager	100	10	8	0.0063	9	7	0.0065	8	6	0.0055
	1000	12	9	0.0192	10	7	0.0158	10	7	0.0150
	5000	13	9	0.0811	11	7	0.0565	12	8	0.0595
	10000	15	9	0.1388	13	7	0.1200	12	6	0.0845
Generalized PSC1	100	22	16	0.0122	20	14	0.0117	20	14	0.0122
	1000	21	15	0.0277	19	13	0.0250	18	12	0.0234
	5000	20	14	0.0835	18	12	0.0709	18	12	0.0753
	10000	19	13	0.1761	18	12	0.1530	17	11	0.1516
Extended Tridiagonal-2	100	46	44	0.0283	46	44	0.0303	47	44	0.0305
	1000	14	12	0.0274	13	11	0.0285	14	12	0.0310
	5000	7	5	0.0283	8	6	0.0389	7	5	0.0368
	10000	7	5	0.0643	7	5	0.0750	6	4	0.0520
Extended Three Exponential Terms	100	9	6	0.0038	10	7	0.0045	9	6	0.0042
	1000	8	5	0.0164	10	7	0.0273	9	6	0.0176
	5000	8	5	0.0461	9	6	0.0522	8	5	0.0514
	10000	8	5	0.0635	9	6	0.0693	8	5	0.0598

应用 Dolan 和 Moré<sup>[16]</sup> 的剖面分析法, 进一步借助 MATLAB 软件分别比较了三个算法的函数估值次数、迭代次数 (见图 1) 以及 CPU 时间 (见图 2). 由图显见, 算法 GDQN1 和 GDQN2 的数值表现总体优于算法 DQN, 其中算法 GDQN2 的表现最佳. 初步的数值实验表明新的对角校正方法较文献 [8] 的校正方法有了较大的改进, 提高了对角拟牛顿算法的效率.

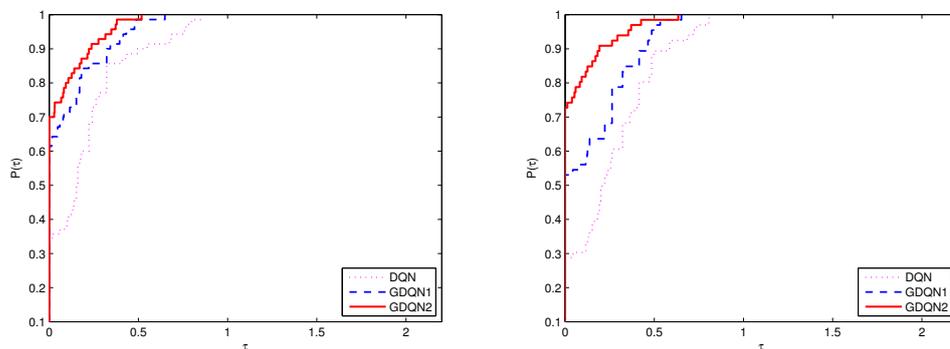


图 1: 函数估值次数 (左) 和迭代次数 (右)

## 5 结论

本文借助弱拟牛顿方程, 对文献 [8] 的对角稀疏拟牛顿法进行了推广与改进, 提出了广义对角拟牛顿法, 主要有以下几方面特点:

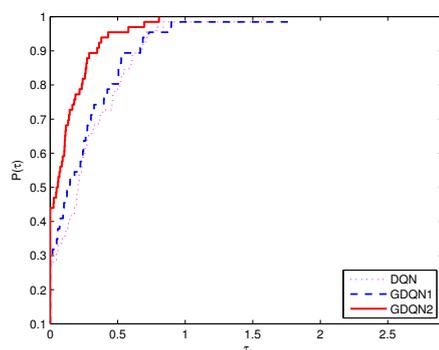


图 2: CPU 时间

(1) 为获取更良好的 Hesse 阵逆矩阵的近似, 合理利用弱拟牛顿方程, 推导出了一类新的 Hesse 阵逆矩阵的对角校正公式;

(2) 用对角矩阵逼近 Hesse 阵的逆矩阵, 在计算过程中不用存储和计算矩阵, 有利于大型问题的求解;

(3) 引进了新的非单调技术提高算法的效率.

### 参 考 文 献

- [1] Grippo L, Lampariello F, Lucidi S. A nonmonotone line search technique for Newton's method[J]. SIAM J. Numer. Anal., 1986, 23(4): 707-716.
- [2] Deng N Y, Xiao Y, Zhou F J. Nonmonotonic trust region algorithm[J]. J. Optim. Theory Appl., 1993, 76(2): 259-285.
- [3] Zhang H C, Hager W W. A nonmonotone line search technique and its application to unconstrained optimization[J]. SIAM J. Optim., 2004, 14(4): 1043-1056.
- [4] Mo J, Liu C, Yan S. A nonmonotone trust region method based on nonincreasing technique of weighted average of the successive function values[J]. J. Comput. Appl. Math., 2007, 209(1): 97-108.
- [5] Gu N, Mo J. Incorporating nonmonotone strategies into the trust region method for unconstrained optimization[J]. Comput. Math. Appl., 2008, 55(9): 2158-2172.
- [6] Barzilai J, Borwein J M. Two point step size gradient method[J]. IMA J. Numer. Anal., 1988, 8(1): 141-148.
- [7] Raydan M. The Barzilai and Borwein gradient method for large scale unconstrained minimization problem[J]. SIAM J. Optim., 1997, 7(1): 26-33.
- [8] 时贞军, 孙国. 无约束优化问题的对角稀疏拟牛顿算法 [J]. 系统科学与数学, 2006, 26(1): 101-112.
- [9] 孙清滢, 刘丽敏, 王宣战. 修正 Grippo 非单调线搜索规则的新对角稀疏拟牛顿算法 [J]. 高等学校计算数学学报, 2011, 33(3): 203-214.
- [10] 孙清滢, 郑艳梅. 大步长非单调线搜索规则的 Lampariello 修正稀疏拟牛顿算法的全局收敛性 [J]. 数学进展, 2008, 37(3): 311-320.
- [11] Yuan Y. A modified BFGS algorithm for unconstrained optimization[J]. IMA J. Numer. Anal., 1991, 11(3): 325-332.

- [12] Zhang J, Xu C. Properties and numerical performance of quasi-Newton methods with modifies quasi-Newton equations[J]. J. Comput. Appl. Math., 2001, 137(2): 269–278.
- [13] Wei Z X, Li G Y, Qi L Q. New quasi-Newton methods for unconstrained optimization[J]. Appl. Math. Comput., 2006, 175(2): 1156–1188.
- [14] Zhu M, Nazareth J L, Wolkowicz H. The quasi-Cauchy and diagonal updating[J]. SIAM J. Optim., 1999, 9(4): 1192–1204.
- [15] Andrei N. An unconstrained optimization test functions collection[J]. Adv. Model. Optim., 2008, 10(1): 147–161.
- [16] Dolan E D, Moré J J. Benchmarking optimization software with performance profiles[J]. Math. Program., 2002, 91(2): 201–213.

## NONMONOTONE GENERALIZED DIAGONAL QUASI-NEWTON ALGORITHM

ZHOU Qun-yan<sup>1</sup>, HANG Dan<sup>2</sup>

(1. School of Mathematics and Physics, Jiangsu University of Technology, Changzhou 213001, China )

(2. Department of Basic Courses, Air Force Logistics College, Xuzhou 221000, China)

**Abstract:** In this paper, the method for solving unconstrained optimization is studied. By using the new diagonal quasi-Newton update and nonmonotone technique, a nonmonotone generalized diagonal quasi-Newton algorithm is proposed. The new method needs less memory capacitance and computational complexity. It is very effective and attractive for large scale unconstrained problems and generalizes the results of reference [8].

**Keywords:** weak quasi-Newton equation; diagonal updating; nonmonotone technique; global convergence; numerical experiment

**2010 MR Subject Classification:** 65K05; 90C30