一种 DHF 混合共轭梯度方法的收敛性证明

数学强基 22 DBQDSS
学号：2221145140

2024 年 5 月 17 日

目录

1 问题背景

共轭梯度法（Conjugate Gradient Method，简称 CG）的发展历史可以追溯至 1952 年，当时 Hestenes 和 Stiefel 首次提出了这种方法，并在论文中进行了详细的解释。这种方法主要用于解线性方程组，特别是在高维空间上的最小化问题。

该方法最初是基于对最速下降法的改进。最速下降法在每一步迭代中都沿着目标函数的负梯度方向前进，但这种方法在接近最优解时收敛速度较慢。共轭梯度法则通过引入共轭方向的概念，使得在迭代过程中搜索方向之间保持一定的正交性，从而加速收敛；同时，它仅需利用一阶导数信息，但克服了最速下降法收敛慢的缺点，又避免了牛顿法需要存储和计算 Hesse 矩阵并求逆的缺点。这使得共轭梯度法成为解决大型线性方程组以及大型非线性最优化问题的有效算法之一。

在 1960 年代，Fletcher 和 Reeves 等人将共轭梯度法扩展到非线性最优化问题。他们提出的非线性共轭梯度法通过计算目标函数的梯度和利用线搜索技术来确定搜索步长，从而逐步逼近最优解。

在过去的几十年里，共轭梯度法的研究得到了广泛的关注。科学家们不断地提出了新的算法和优化方法，以提高共轭梯度法的收敛速度和稳定性。这些研究使得共轭梯度法在各种应用场景中的性能得到了显著的提高。

现在，共轭梯度法已经广泛地应用于实际问题中，包括计算机视觉、机器学习、社交网络分析等领域。例如，在机器学习任务中，共轭梯度法（特别是其随机梯度下降版本）表现出色，可以在大规模数据集上实现并行计算和高效的内存使用。

总的来说，共轭梯度法的发展历史是一个不断演进和优化的过程，其应用范围也在不断扩大。

下面的 DHF 方法就是一种混合共轭梯度法。

2 相关证明

2.1 题目描述

共轭梯度法 (Conjugate Gradient , CG)，最初是为了求解线性系统和无约束最小化而提出的。该方法是科学家、工程师和数学家解决优化问题的

极佳选择。考虑以下问题

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

其中 $f:\mathbb{R}^n\to\mathbb{R}$ 是光滑的非线性函数。共轭梯度法的迭代格式为：

$$\begin{cases} x_0\in\mathbb{R}^n \\ x_{k+1}=x_k+\alpha_k d_k \end{cases} \tag{2}$$

其中 $\alpha_k>0$ 是搜索方向。这里考虑强 Wolfe 准则下的非精确线性搜索，即 α_k 需满足：

$$f(x_k+\alpha_k d_k)\leqslant f(x_k)+\delta\alpha_k g_k^T d_k \tag{3}$$

$$|g(x_k+\alpha_k d_k)^T d_k|\leqslant -\sigma g_k d_k \tag{4}$$

其中 $0<\delta<\sigma<1$，搜索方向:

$$d_{k+1}=-g_{k+1}+\beta_k d_k \tag{5}$$

其中 β_k 称为共轭梯度参数。对于非线性共轭梯度方法，共轭条件由下式给出：

$$d_{k+1}^T y_k=0 \tag{6}$$

Perry (1978) 通过利用拟牛顿法的割线条件 $B_{k+1}s_k=y_k$ 和由 $B_{k+1}d_{k+1}=-g_{k+1}$ 给出的拟牛顿搜索方向扩展了 (2) 式中的结果，其中 B_{k+1} 是 Hessen 近似对称正定矩阵，如下：

$$d_{k+1}^T y_k=-g_{k+1}^T s_k \tag{7}$$

然而，实际中得数值计算通常采用不精确线性搜索，即 $-g_{k+1}^T s_k\neq 0$ ，因此 Dai 和 Liao (2001) 用扩展共轭条件改写了 (3) 式为：

$$d_{k+1}^T=-t g_{k+1}^T s_k,\quad t\geqslant 0 \tag{8}$$

令 $\|\cdot\|$ 表示欧式范数，记 $s_k=x_{k+1}-x_k$，$y_k=g_{k+1}-g_k$，则 Hestenes-Stiefel 公式和 Fletcher-Reeves 公式为：

$$\beta_k^{HS}=\frac{g_{k+1}^T y_k}{d_k^T y_k} \tag{9}$$

$$\beta_k^{FR}=\frac{\|g_{k+1}\|^2}{\|g_k\|^2} \tag{10}$$

下面我们考虑 HS 和 FR 共轭梯度方法的凸组合：

$$\beta_k^{DHF} = (1-\theta_k)\beta_k^{HS} + \theta_k \beta_k^{FR} \tag{11}$$

也即

$$\beta_k^{DHF} = (1-\theta_k)\frac{g_{k+1}^T y_k}{d_k^T y_k} + \theta_k \frac{\|g_{k+1}\|^2}{\|g_k\|^2} \tag{12}$$

其中 $\theta_k \in [0,1]$ 称为杂交标量参数。若 $\theta_k \leqslant 0$，则令 $\theta_k = 0$，那么 $\beta_k^{DHF} = \beta_k^{HS}$；若 $\theta_k \geqslant 1$，则令 $\theta_k = 1$，那么 $\beta_k^{DHF} = \beta_k^{FR}$. 因此下降方向为：

$$d_{k+1} = -g_{k+1} + \left((1-\theta_k)\frac{g_{k+1}^T y_k}{d_k^T y_k} + \theta_k \frac{\|g_{k+1}\|^2}{\|g_k\|^2}\right) d_k \tag{13}$$

$$d_{k+1}^T y_k = -g_{k+1}^T y_k + \left((1-\theta_k)\frac{g_{k+1}^T y_k}{d_k^T y_k} + \theta_k \frac{\|g_{k+1}\|^2}{\|g_k\|^2}\right) d_k^T y_k \tag{14}$$

考虑 $d_{k+1} = -\nabla^2 f(x_{k+1})^{-1} g_{k+1}$，有 $d_{k+1}^T y_k = -\nabla^2 f(x_{k+1})^{-1} g_{k+1}^T y_k$，联立 (7) 式解得：

$$\theta_k = \frac{(-s_k^T g_{k+1})\|g_{k+1}\|^2}{(y_k^T s_k)\|g_{k+1}\|^2 - (g_{k+1}^T y_k)\|g_k\|^2} \tag{15}$$

于是 d_{k+1} 满足牛顿方向。再将 Dai-Liao 共轭条件 (4) 代入 (11) 中，得到另一个杂交参数：

$$\theta_k = \frac{(-t s_k^T g_{k+1})\|g_{k+1}\|^2}{(y_k^T s_k)\|g_{k+1}\|^2 - (g_{k+1}^T y_k)\|g_k\|^2} \tag{16}$$

其中我们根据 Babaie-Kafaki 和 Ghanbari (2015) 以及 Andrei (2017) 中获得的最优选择来设置参数 t 为：

$$t^* = \frac{s_k^T y_k}{s_k^T s_k} \tag{17}$$

2.2 第一问

考虑具有搜索方向 (5) 和由 (12) 生成的的 DHF 共轭梯度算法，证明以下充分下降条件成立：

$$d_k^T g_k \leqslant -c\|g_k\|^2, \quad \forall k \geqslant 0$$

其中 c 是一个正的常数。

证明：

由题意知，若存在这样的 c，则有：

$$\begin{aligned} d_k^T g_k \leqslant -c\left\|g_k\right\|^2 &\Leftrightarrow (-g_{k+1}+\beta_k d_k)^T g_{k+1} = -c\left\|g_{k+1}\right\| \\ &\Leftrightarrow \beta_k d_k^T g_{k+1} \leqslant (1-c)\left\|g_{k+1}\right\|^2 \end{aligned}$$

注意到，

$$\beta_k = \beta_k^{DHF} = (1-\theta_k)\beta_k^{HS} + \theta_k \beta_k^{FR}$$

这个式子是关于 θ_k 的一次函数，于是左侧式子最大值必定在端点处取到，也即

$$\max\{\beta_k d_k^T g_{k+1}\} = \max\{\beta_k^{HS} d_k^T g_{k+1}, \beta_k^{FR} d_k^T g_{k+1}\}$$

于是，原命题转化为去证明对于 HS 公式与 FR 公式分别有符合题意的正数 c 存在。

情形一：HS 公式

考虑 Powell 重启准则 [1]，有

$$|g_{k+1}^T g_k| > \xi \left\|g_{k+1}\right\|^2$$

其中，ξ 是某一正常数。

若上述准则成立，则命题已成立；否则有：

$$|g_{k+1}^T g_k| \leqslant \xi \left\|g_{k+1}\right\|^2 \tag{18}$$

其中，ξ 是某一正常数。

按照 HS 公式的定义，有：

$$d_{k+1}^T g_{k+1} = -g_{k+1}^T g_{k+1} + \frac{g_{k+1}^T y_k}{d_k^T y_k} d_k^T g_{k+1} \leqslant -\left\|g_{k+1}\right\|^2 + \frac{|d_k^T g_{k+1}|}{|d_k^T y_k|} |g_{k+1}^T y_k|$$

下面讨论后一项的上界，最后一步用到 (18) 式，

$$\begin{aligned}
|d_k^T g_{k+1}| &\leqslant \sigma |d_k^T g_k| \\
|d_k^T y_k| &= |d_k^T g_{k+1} - d_k^T g_k| \geqslant \left| |d_k^T g_{k+1}| - |d_k^T g_k| \right| \\
&= |1-\sigma| |d_k^T g_k| = (1-\sigma)|d_k^T g_k| \\
|g_{k+1}^T y_k| &= |g_{k+1}^T (g_{k+1} - g_k)| \leqslant \|g_{k+1}\|^2 + |g_{k+1}^T g_k| \\
&\leqslant (1+\xi) \|g_{k+1}\|^2
\end{aligned}$$

于是有，

$$\begin{aligned}
d_{k+1}^T g_{k+1} &\leqslant - \|g_{k+1}\|^2 + \frac{|d_k^T g_{k+1}|}{|d_k^T y_k|} |g_{k+1}^T y_k| \\
&\leqslant - \|g_{k+1}\|^2 + \frac{\sigma(1+\xi)}{1-\sigma} \|g_{k+1}\|^2 \\
&= \frac{(2+\xi)\sigma - 1}{1-\sigma} \|g_{k+1}\|^2
\end{aligned}$$

取满足下面条件的 c, ξ, σ 即可：

$$c = \frac{1-(2+\xi)\sigma}{1-\sigma} > 0, \quad (2+\xi)\sigma < 1$$

情形二：FR 公式

按照 FR 公式的定义，有：

$$\frac{\|g_{k+1}\|^2}{\|g_k\|^2} d_k^T g_{k+1} \leqslant (1-c) \|g_{k+1}\|^2 \Leftrightarrow \frac{d_k^T g_{k+1}}{\|g_k\|^2} \leqslant 1-c$$

我们使用 (4) 式与 Al-Baali [2] 给出的不等式估计出左式的上界（此时我们需要选定 $\sigma \in (0, \frac{1}{2})$）：

$$\begin{aligned}
\frac{d_k^T g_{k+1}}{\|g_k\|^2} \leqslant \frac{|d_k^T g_{k+1}|}{\|g_k\|^2} &\leqslant \frac{\sigma(-d_k^T g_k)}{\|g_k\|^2} \\
&\leqslant \frac{\sigma(1-\sigma^{k+1})}{1-\sigma} \\
&< \frac{\sigma}{1-\sigma}
\end{aligned}$$

令

$$\frac{\sigma}{1-\sigma} = 1 - c$$

解出：

$$c = \frac{1-2\sigma}{1-\sigma} > 0, \quad 0 < \sigma < \frac{1}{2}$$

这样的 c 符合题意，于是原命题在 FR 公式的情形下亦成立。
原命题得证！ □

2.3 第二问

假设 (i)： 水平集 $S = x \in \mathbb{R} : f(x) \leqslant f(x_0)$ 是有界的，即存在一个正的常数 B 使得 $\|x\| \leqslant B, \forall x \in S$。
假设 (ii)： 在 S 的一个邻域 N 内，目标函数 f 连续可微，并且其梯度 $g(x)$ 在 N 上是 Lipschitz 连续的，即存在一个常数 $L > 0$ 使得：

$$\|f(x) - f(y)\| \leqslant L\,\|x - y\|\,, \quad \forall x, y \in N$$

在的假设 (i) 和假设 (ii) 下，存在一个常数 Γ 使得：

$$\|g(x)\| \leqslant \Gamma, \quad \forall x \in S$$

引理 1： 令假设 (i) 和假设 (ii) 成立，考虑由迭代格式 (2) 和下降方向 (5) 构成的共轭梯度法，其中满足强 Wolfe 准则 (3) 和 (4)，若

$$\sum_k \frac{1}{\|d_k\|^2} = \infty$$

则有

$$\lim_{x\to 0} \inf \|g_k\| = 0$$

已知上述假设 (i),(ii) 与引理 1 均成立，证明 DHF 算法的收敛性：对某个 k 有 $g_k = 0$ 或者

$$\lim_{k\to\infty} \inf \|g_k\| = 0$$

注：引理 1 的证明可见于参考文献 [3]。

证明：

首先证明一个引理：

引理 2：假设 d_k 是下降方向，梯度 $g(x)$ 满足 Lipschitz 条件：$\|g(x_1) - g(x_2)\| \leqslant L\|x_1 - x_2\|, \forall x_1, x_2 \in N$，而且还满足强 Wolfe 条件，则步长因子 α_k 有：

$$\alpha_k \geqslant \frac{1-\sigma}{L}\frac{|g_k^T d_k|}{\|d_k\|^2}$$

引理 2 的证明: 由强 Wolfe 准则：

$$\begin{aligned}
&\sigma g_k^T d_k \leqslant g_{k+1}^T d_k \\
&\Rightarrow (\sigma - 1) g_k^T d_k \leqslant (g_{k+1} - g_k)^T d_k \\
\text{（Lipschitz 条件）} \quad &\Rightarrow (\sigma - 1) g_k^T d_k \leqslant L\|d_k\|^2 \alpha_k \\
&\Rightarrow \alpha_k \geqslant \frac{1-\sigma}{L}\frac{|g_k^T d_k|}{\|d_k\|^2}
\end{aligned}$$

回到原题，假设 $g_k \neq 0, \forall k > 0$，下证：$\liminf_{k\to 0}\|g_k\| = 0$ 成立。使用反证法，假设 $\liminf_{k\to 0}\|g_k\| > 0$，则存在 $\epsilon > 0$，使得当 k 充分大时，有 $\|g_k\| > \epsilon$。

由假设 (i)，

$$\|s_k\| = \|x_{k+1} - x_k\| \leqslant \|x_{k+1}\| + \|x_k\| \leqslant 2B$$

则可估计 d_{k+1} 的上界，

$$\begin{aligned}
\|d_{k+1}\| &= \left\|-g_{k+1} + \beta_k^{DHF} d_k\right\| \\
&= \left\|-g_{k+1} + [(1-\theta_k)\beta_k^{HS} + \theta_k \beta_k^{FR}] d_k\right\| \\
&= \|-g_{k+1}\| + (|(1-\theta_k)||\beta_k^{HS}| + |\theta_k||\beta_k^{FR}|)\|d_k\| \\
&< \|-g_{k+1}\| + (|\beta_k^{HS}| + |\beta_k^{FR}|)\|d_k\| \\
&\leqslant \Gamma + (|\beta_k^{HS}| + |\beta_k^{FR}|)\|d_k\|
\end{aligned}$$

下面分别讨论 $|\beta_k^{HS}|$、$|\beta_k^{FR}|$ 与 $\|d_k\|$ 的上界，

$$
\begin{aligned}
|\beta_k^{HS}| &= \left|\frac{g_{k+1}^T y_k}{d_k^T y_k}\right| \leqslant \frac{\|g_{k+1}\| \, \|y_k\|}{|d_k^T y_k|} \\
&\leqslant \frac{\Gamma L \|s_k\|}{c \|g_k\|^2} \leqslant \frac{2\Gamma L B}{c\epsilon^2} \\
|\beta_k^{FR}| &= \frac{\|g_{k+1}\|^2}{\|g_k\|^2} \leqslant \frac{\Gamma^2}{\epsilon^2} \\
\|d_k\| &= \frac{\|s_k\|}{\alpha_k} \leqslant \frac{2B}{T} \quad (\text{由引理 } 2)
\end{aligned}
$$

记 $A = \frac{2\Gamma LB}{c\epsilon^2} + \frac{\Gamma^2}{\epsilon^2}$，$M = \frac{2B}{T}$，于是有，

$$
\|d_{k+1}\| \leqslant \|g_{k+1}\| + |\beta_k| \, \|d_k\| \leqslant \Gamma + AM
$$

于是，

$$
\sum_{k \geqslant 0} \frac{1}{\|d_k\|^2} = \infty
$$

由引理 1 可得，

$$
\lim_{k \to \infty} \inf \|g_k\| = 0
$$

原命题得证！ $\square$

参考文献

[1] Hager W W, Zhang H. A survey of nonlinear conjugate gradient methods[J]. Pacific journal of Optimization, 2006, 2(1): 35-58.

[2] M. Al-Baali, Descent property and global convergence of the Fletcher-Reeves method with inexact line search, IMA J. Numer. Anal., 5 (1985), pp. 121–124.

[3] Dai Y H. Nonlinear conjugate gradient methods[J]. Wiley Encyclopedia of Operations Research and Management Science, 2010.