广西科学 Guangxi Sciences 2011,18(1):39～43

梯度相关条件下求解非线性规划问题算法的收敛性*

Convergence of Some Algorithms under the Condition of Gradient Correlation

陈进来,高苏銮,韦增欣

CHEN Jin-lai,GAO Su-luan,WEI Zeng-xin

(广西大学数学与信息科学学院,广西南宁 530004)

(College of Mathematics and Information Science, Guangxi University, Nanning, Guangxi, 530004,China)

摘要:证明在梯度相关的条件下,一般的求解非线性规划问题算法的收敛性定理在 $\theta_k = \langle -\nabla f(x_k), d_k\rangle$ 无需满足其它任何条件的前提下仍然成立.

关键词:线搜索 梯度相关 收敛性

中图法分类号:O221.2 **文献标识码**:A **文章编号**:1005-9164(2011)01-0039-05

Abstract: We prove that the convergence theorems of general algorithms for solving nonlinear programming problems hold under the condition of gradient correlation, meanwhile $\theta_k = \langle -\nabla f(x_k), d_k\rangle$ does not satisfy any condition.

Key words: line search, gradient correlation, convergence

在一般的求解无约束优化问题的算法中,最关键的环节是求下降方向 d_k 和迭代步长 α_k. 求迭代步长常用的方法有精确线搜索和非精确线搜索. 例如,

精确一维搜索[1]: α_k 满足

$$f(x_k + \alpha_k) = \min_{\alpha \geqslant 0} f(x_k + \alpha d_k). \tag{1}$$

Wolfe-Powell 线搜索[1](WP 准则): α_k 满足

$$f(x_k + \alpha_k d_k) - f(x_k) \leqslant \sigma\alpha_k \nabla f(x_k)^T d_k, \tag{2}$$

$$\nabla f(x_k + \alpha_k d_k)^T d_k \geqslant \beta \nabla f(x_k)^T d_k, \tag{3}$$

其中 $\sigma \in (0, \frac{1}{2})$, $\beta \in (\sigma, 1)$.

Armijo-Goldstein 线搜索[1](AG 准则): α_k 满足

$$(1-\sigma)\alpha_k \nabla f(x_k)^T d_k \leqslant f(x_k + \alpha_k d_k) - f(x_k) \leqslant \sigma\alpha_k \nabla f(x_k)^T d_k, \tag{4}$$

其中 $\sigma \in (0, \frac{1}{2})$.

Armijo 线搜索[2]:令 $\alpha_k = \beta^m s$,其中 m 是满足下式的最小非负整数

$$f(x_k + \beta^m s d_k) - f(x_k) \leqslant s\beta^m \sigma \nabla f(x_k)^T d_k. \tag{5}$$

其中 $\sigma \in (0,1)$, $\beta \in (0,1)$, $s > 0$.

文献[3]用 $f_A(x) = f(x) + (x - x_k)^T A_k (x - x_k)$($A_k$ 是正定矩阵)代替上述线搜索中的 $f(x)$ 得到几种新的线搜索:

修正的 Wolfe-Powell 线搜索(MWP 准则): α_k 满足

$$f(x_k + \alpha_k d_k) + \frac{1}{2}\alpha_k^2 d_k^T A_k d_k - f(x_k) \leqslant \sigma\alpha_k \nabla f(x_k)^T d_k, \tag{6}$$

$$\nabla f(x_k + \alpha_k d_k)^T d_k + \alpha_k d_k^T A_k d_k \geqslant \beta \nabla f(x_k)^T d_k. \tag{7}$$

修正的 Armijo-Goldstein 线搜索(MAG 线准则): α_k 满足

$$(1-\sigma)\alpha_k \nabla f(x_k)^T d_k \leqslant f(x_k + \alpha_k d_k) + \frac{1}{2}\alpha_k^2 d_k^T A_k d_k - f(x_k) \leqslant \sigma\alpha_k \nabla f(x_k)^T d_k. \tag{8}$$

修正的 Armijo 线搜索(MA 准则):令 $\alpha_k = \beta^m s$,其中 m 是满足下式的最小非负整数

$$f(x_k + \beta^m s d_k) + \frac{1}{2}\alpha_k^2 d_k^T A_k d_k - f(x_k) \leqslant s\beta^m \sigma \nabla f(x_k)^T d_k.$$

已有学者对这些不同线搜索所对应算法的收敛

收稿日期:2010-06-09

作者简介:陈进来(1984-),男,硕士研究生,主要从事优化与管理,交通与物流管理。

*国家自然科学基金项目(10761001)资助。

性进行研究[2~9]，得到了算法的收敛速度和收敛性定理. 文献[3]证明修正线搜索的存在性及其所对应的算法在 $\theta=\langle-\nabla f(x_k),d_k\rangle$ 满足 $\langle f(x_k),d_k\rangle$ 大于 0 小于 90 度的收敛性. 文献[5]讨论常见的 3 种非精确线搜索在梯度相关条件下的一些性质. 本文在前人所得收敛性定理的基础上，结合梯度相关给出上述线搜索所对应算法的新收敛性定理.

1 WP 线搜索和 MWP 线搜索的简单性质

对于求解一般的无约束优化问题 $\min f(x), x\in R^n$，算法[1]，我们规定以下假定成立：

(1)对任意的 k 及 x_k，搜索方向 d_k 为下降方向；

(2)对任意的 k，A_k 是正定对称矩阵，且存在 $C_2\geqslant C_1>0$ 使得 $\forall x\in R^n$，$C_1x^Tx\leqslant x^TA_kx\leqslant C_2x^Tx$.

若求步长因子 α_k 时采用 WP 准则，则有

定理 1.1[1] 设函数 $f(x)$ 连续可微，$\nabla f(x)$ 满足 Lipschitz 条件

$$\|\nabla f(y)-\nabla f(x)\|\leqslant M\|y-x\|.$$

如果 $f(x_k+\alpha_kd_k)$ 有下界，$\sigma>0$，令 $\theta_k=\langle d_k,-\nabla f(x_k)\rangle$，则对满足 WP 线搜索的任何 α_k 均有

$$f(x_k)-f(x_k+\alpha_kd_k)\geqslant\frac{\sigma(1-\beta)}{M}\cdot\|\nabla f(x_k)\|^2\cos^2\theta_k. \tag{9}$$

类似地，若求步长因子 α_k 时采用 MWP 准则，有

定理 1.2 设函数 $f(x)$ 连续可微，$\nabla f(x)$ 满足 Lipschitz 条件

$$\|\nabla f(y)-\nabla f(x)\|\leqslant M\|y-x\|.$$

如果 $f(x_k+\alpha_kd_k)$ 有下界，$\sigma>0$，令 $\theta_k=\langle d_k,-\nabla f(x_k)\rangle$，则对满足 MWP 线搜索的任何 α_k 均有

$$f(x_k)-f(x_k+\alpha_kd_k)\geqslant\frac{\sigma(1-\beta)}{M}\cdot\|\nabla f(x_k)\|^2\cos^2\theta_k-\sigma\|\nabla f(x_k)\|\cdot\cos\theta_k\frac{\alpha_kd_k^TA_kd_k}{M\|d_k\|}.$$

证明 由 Lipschitz 条件可知

$$d_k^T[\nabla f(x_k+\alpha_kd_k)-\nabla f(x_k)]\leqslant\|d_k\|\cdot\|\nabla f(x_k+\alpha_kd_k)-\nabla f(x_k)\|\leqslant M\alpha_k\|d_k\|.$$

由于 α_k 满足(7)式，故有

$$d_k^T[\nabla f(x_k+\alpha_kd_k)-\nabla f(x_k)]\geqslant-(1-\beta)d_k^T\cdot\nabla f(x_k)-\alpha_kd_k^TA_kd_k.$$

所以

$$M\alpha_k\|d_k\|^2\geqslant-(1-\beta)d_k^T\nabla f(x_k)-\alpha_kd_k^TA_k\cdot d_k=(1-\beta)\cos\theta_k\|d_k\|\|\nabla f(x_k)\|-\alpha_kd_k^TA_kd_k.$$

故

$$\alpha_k\|d_k\|\geqslant\frac{1-\beta}{M}\cos\theta_k\|\nabla f(x_k)\|-\frac{\alpha_kd_k^TA_kd_k}{M\|d_k\|},$$

又因为 α_k 满足(6)式，可知

$$f(x_k)-f(x_k+\alpha_kd_k)\geqslant-\sigma\alpha_k\nabla f(x_k)^Td_k+\frac{1}{2}\alpha_k^2d_k^TA_kd_k\geqslant-\sigma\alpha_k\nabla f(x_k)^Td_k=\sigma\alpha_k\|d_k\|\cos\theta_k\|\nabla f(x_k)\|\geqslant\sigma\cos\theta_k\|\nabla f(x_k)\|(\frac{1-\beta}{M}\cos\theta_k\|\nabla f(x_k)\|-\frac{\alpha_kd_k^TA_kd_k}{M\|d_k\|})=\frac{\sigma(1-\beta)}{M}\cos^2\theta_k\|\nabla f(x_k)\|^2-\sigma\cos\theta_k\|\nabla f(x_k)\|\frac{\alpha_kd_k^TA_kd_k}{M\|d_k\|}.$$

2 梯度相关性质

定义 2.1[2] 假设 $f(x)$ 连续可微，且在水平集 $L(x_0)=\{x\mid f(x)\leqslant f(x_0)\}$ 上有界，如果 $x^*\in R^n$ 满足 $\nabla f(x^*)=0$，则称 x^* 为 $f(x)$ 的一个稳定点.

定义 2.2[2] 假设 $\{x_k\}$ 是由算法 $x_{k+1}=x_k+\alpha_kd_k$ 产生，若对任何 $\{x_k\}$ 的子序列 $\{x_k\}_\Gamma\to x^*$，x^* 为一个非稳定点，$\{d_k\}_\Gamma$ 有界且满足

$$\lim_{k\to\infty,k\in\Gamma}\sup\nabla f(x_k)^Td_k<0, \tag{10}$$

则称 $\{d_k\}$ 与 $\{x_k\}$ 是梯度相关的.

性质 2.1 $f(x)$ 为连续可微函数，$\{x_k\}$ 是由算法 $x_{k+1}=x_k+\alpha_kd_k$ 产生的点列，如果算法产生的下降方向 $\{d_k\}$ 与点列 $\{x_k\}$ 梯度相关，且存在常数 $\alpha>0$ 使得 $\|\nabla f(x_k)\|\geqslant\alpha$. 令 $\theta_k=\langle d_k,-\nabla f(x_k)\rangle$，则 $\cos\theta_k$ 不趋近于 0.

证明 因为 $\{d_k\}$ 与 $\{x_k\}$ 是梯度相关，所以

$$\lim_{k\to\infty}\sup\nabla f(x_k)^Td_k<0.$$

由于 $\cos\theta_k=\dfrac{-\nabla f(x_k)^Td_k}{\|\nabla f(x_k)\|\cdot\|d_k\|}$，故 $\nabla f(x_k)^Td_k=-\cos\theta_k\|\nabla f(x_k)\|\cdot\|d_k\|$. 由 $\|\nabla f(x_k)\|\geqslant\alpha$，可知

$$-\alpha\cos\theta_k\|d_k\|\leqslant-\cos\theta_k\|\nabla f(x_k)\|\cdot\|d_k\|=\nabla f(x_k)^Td_k.$$

所以

$$\lim_{k\to\infty}-\alpha\cos\theta_k\|d_k\|\leqslant\limsup_{k\to\infty}-\cos\theta_k\cdot\|\nabla f(x_k)\|\cdot\|d_k\|=\limsup_{k\to\infty}\nabla f(x_k)^Td_k<0.$$

若 $\cos\theta_k\to0$，由 $\|d_k\|$ 有界可知 $\lim\limits_{k\to\infty}-\alpha\cos\theta_k\|d_k\|=0$. 这与 $\lim\limits_{k\to\infty}-\alpha\cos\theta_k\|d_k\|<0$ 矛盾，故 $\cos\theta_k$ 不趋近于 0.

3 算法的收敛性

为了方面，在以下讨论中记 $g_k=\nabla f(x_k)$.

定理 3.1 设在算法中采用精确一维搜索求步长 α_k，且算法产生的下降方向 $\{d_k\}$ 与点列 $\{x_k\}$ 梯度相关. 若 $\nabla f(x)$ 在水平集 $L=\{x\in R^n \mid f(x)\leqslant f(x_0)\}$ 存在且一致连续，则要么对某个 k 有 $g_k=0$，要么有 $f(x_k)\to-\infty$，要么有 $g_k\to 0$.

证明 假设对所有的 k 有 $g_k\neq 0$，且 $f(x_k)$ 有下界. 由于 $\{f(x_k)\}$ 单调递减，故 $\{f(x_k)\}$ 的极限存在，因而

$$f(x_k)-f(x_{k+1})\to 0. \tag{11}$$

假定 $g_k\to 0$ 不成立，则存在常数 $\varepsilon>0$ 和一个子列使得 $\|g_k\|\geqslant\varepsilon$. 而 $\{d_k\}$ 与 $\{x_k\}$ 是梯度相关的，由性质 2.1 可知，$\cos\theta_k\to 0$ 不成立，则存在 $\varepsilon_0>0$ 和一个子列使得 $|\cos\theta_k|\geqslant\varepsilon_0$，从而

$$\frac{-g_k^Td_k}{\|d_k\|}=\|g_k\|\cdot\cos\theta_k\geqslant\varepsilon\cos\theta_k=\varepsilon\,|\cos\theta_k|\geqslant\varepsilon\cdot\varepsilon_0\triangleq\varepsilon_1, \tag{12}$$

又

$$f(x_k+\alpha d_k)=f(x_k)+\alpha g(\xi_k)^Td_k=f(x_k)+\alpha g_k^Td_k+\alpha[g(\xi_k)-g_k]^Td_k\leqslant f(x_k)+\alpha\|d_k\|\left(\frac{g_k^Td_k}{\|d_k\|}+\|g(\xi_k)-g_k\|\right), \tag{13}$$

其中 ξ_k 在 x_k 与 $x_k+\alpha d_k$ 之间. 由于 $\nabla f(x)$ 在水平集 L 上一致连续，故存在 $\delta>0$，使得当 $0\leqslant\alpha\|d_k\|\leqslant\delta$ 时，

$$\|g(\xi_k)-g_k\|\leqslant\frac{1}{2}\varepsilon_1. \tag{14}$$

所以由(12)～(14)式可得，

$$f\left(x_k+\delta\frac{d_k}{\|d_k\|}\right)\leqslant f(x_k)+\delta\left(\frac{g_k^Td_k}{\|d_k\|}+\frac{1}{2}\varepsilon_1\right)\leqslant f(x_k)-\frac{1}{2}\delta\varepsilon_1.$$

从而 $f(x_{k+1})\leqslant f\left(x_k+\delta\frac{d_k}{\|d_k\|}\right)\leqslant f(x_k)-\frac{1}{2}\delta\varepsilon_1$，即 $\frac{1}{2}\delta\varepsilon\leqslant f(x_k)-f(x_{k+1})$，这与(11)式矛盾. 从而 $g_k\to 0$.

定理 3.2 设在算法中采用 AG 准则或 WP 准则求步长 α_k，且算法产生的下降方向 $\{d_k\}$ 与点列 $\{x_k\}$ 梯度相关. 若 $\nabla f(x)$ 在水平集 $L=\{x\in R^n \mid f(x)\leqslant f(x_0)\}$ 上存在且一致连续，则要么对某个 k 有 $g_k=0$，要么有 $f(x_k)\to-\infty$，要么有 $g_k\to 0$.

证明 只证 α_k 采用 AG 准则时结论成立. 假设对所有的 k 有 $g_k\neq 0$，且 $f(x_k)$ 有下界. 由于 $\{f(x_k)\}$ 单调递减，故 $\{f(x_k)\}$ 的极限存在，因而 $f(x_k)-f(x_{k+1})\to 0$. 再由(4)式可知，$\sigma\alpha_k\nabla f(x_k)^Td_k\to 0$，且

$$\frac{f(x_k)-f(x_{k+1})}{-\alpha_kg_k^Td_k}\leqslant 1-\sigma, \tag{15}$$

故

$$-\sigma\alpha_k\nabla f(x_k)^Td_k=\sigma\alpha_k\|\nabla f(x_k)\|\,\|d_k\|\cdot\cos\theta_k\to 0. \tag{16}$$

假定 $g_k\to 0$ 不成立，则存在常数 $\varepsilon>0$ 和一个子列使得 $\|g_k\|\geqslant\varepsilon$. 而 $\{d_k\}$ 与 $\{x_k\}$ 是梯度相关的，由性质 2.1 可知，$\cos\theta_k\to 0$ 不成立，则存在 $\varepsilon_0>0$ 和一个子列使得 $|\cos\theta_k|\geqslant\varepsilon_0$. 又 $|\cos\theta_k|\leqslant 1$，由(16)式可知 $\alpha_k\|d_k\|\to 0$. 令 $s_k=\alpha_kd_k$，则 $\|s_k\|\to 0$，且

$$-g_k^Ts_k=\|g_k\|\cdot\|s_k\|\cdot\cos\theta_k\geqslant\varepsilon\cdot\|s_k\|\cos\theta_k=\varepsilon\cdot\|s_k\|\,|\cos\theta_k|\geqslant\varepsilon\cdot\varepsilon_0\cdot\|s_k\|.$$

但由 Taylor 展式可知，$f(x_{k+1})=f(x_k)+g(\xi_k)^Ts_k$，其中 ξ_k 位于 x_k 与 x_{k+1} 之间. 由于 $\nabla f(x)$ 在水平集 L 上一致连续，故当 $s_k\to 0$ 时，$g(\xi_k)\to g_k$. 故 $f(x_{k+1})=f(x_k)+g_k^Ts_k+o(\|s_k\|)$. 由此可得

$$\frac{f(x_k)-f(x_{k+1})}{-g_k^Ts_k}\to 1.$$

这与(15)式矛盾. 故有 $g_k\to 0$.

类似地，当 α_k 采用 WP 准则时定理 3.2 也成立.

定理 3.3 设 $f(x)$ 连续可微，且 $\nabla f(x)$ 满足 Lipischitz 条件

$$\|\nabla f(x)-\nabla f(y)\|\leqslant M\|x-y\|. \tag{17}$$

设在算法中采用 WP 准则求步长 α_k，如果算法产生的下降方向 $\{d_k\}$ 与点列 $\{x_k\}$ 梯度相关，则要么对某个 k 有 $g_k=0$，要么有 $f(x_k)\to-\infty$，要么有 $g_k\to 0$.

证明 反证法. 假设对每一个 k 都有 $g_k\neq 0$，且 $f(x_k)$ 有下界. 由于 $\{f(x_k)\}$ 单调递减，故 $\{f(x_k)\}$ 的极限存在，因而 $f(x_k)-f(x_{k+1})\to 0$. 由(2)式知

$$f(x_k)-f(x_{k+1})\geqslant-\sigma\alpha_kd_k^Tg_k>0. \tag{18}$$

又因为 $f(x_k)-f(x_{k+1})\to 0$，故

$$s_k^Tg_k\to 0, \tag{19}$$

其中 $s_k=\alpha_kd_k$.

假定 $g_k\to 0$ 不成立，则存在常数 $\varepsilon>0$ 和一个子列使得 $\|g_k\|\geqslant\varepsilon$. 而 $\{d_k\}$ 与 $\{x_k\}$ 是梯度相关的，由性质 2.1 可知，$\cos\theta_k\to 0$ 不成立，则存在 $\varepsilon_0>0$ 和一个子列使得 $|\cos\theta_k|\geqslant\varepsilon_0$，从而

$$-g_k^Ts_k=\|g_k\|\cdot\|s_k\|\cdot\cos\theta_k\geqslant\varepsilon\cdot\|s_k\|\cos\theta_k=\varepsilon\cdot\|s_k\|\,|\cos\theta_k|\geqslant\varepsilon\cdot\varepsilon_0\cdot\|s_k\|>0. \tag{20}$$

由(19)式和(20)式可知 $\|s_k\|\to 0$. 又由定理 2.2 可知

$$f(x_k)-f(x_k+\alpha_kd_k)\geqslant\frac{\sigma(1-\beta)}{M}\|\nabla f(x_k)\|^2\cos^2\theta_k\geqslant\frac{\sigma(1-\beta)}{M}\varepsilon^2\varepsilon_0^2.$$

这与 $f(x_k)-f(x_{k+1})\to 0$ 矛盾，从而 $g_k\to 0$.

定理 3.1～3.2 均为常见的线搜索所对应算法的新的收敛性定理，对于修正后的线搜索也具有类似的性质.

定理 3.4 设算法采用 MAG 准则或 MWP 准则求步长 α_k，且算法产生的下降方向 $\{d_k\}$ 与点列 $\{x_k\}$ 梯度相关. 若 $\nabla f(x)$ 在水平集 $L=\{x\in R^n\mid f(x)\leqslant f(x_0)\}$ 上存在且一致连续，则要么对某个 k 有 $g_k=0$，要么有 $f(x_k)\to-\infty$，要么有 $g_k\to 0$.

证明 只证步长 α_k 采用 MAG 准则时结论成立. 假设对所有的 k 有 $g_k\neq 0$ 且 $f(x_k)$ 有下界. 由于 $\{f(x_k)\}$ 单调递减，故 $\{f(x_k)\}$ 的极限存在，因而 $f(x_k)-f(x_{k+1})\to 0$. 因此由 (8) 式可知，$\sigma\alpha_k\nabla f(x_k)^Td_k-\frac{1}{2}\alpha_k^2d_k^TA_kd_k\to 0$，且

$$\frac{f(x_k)-f(x_{k+1})}{-\alpha_kg_k^Td_k}\leqslant(1-\sigma)+\frac{\frac{1}{2}\alpha_k^2d_k^TA_kd_k}{-\alpha_kg_k^Td_k}. \tag{21}$$

令 $s_k=\alpha_kd_k$，则

$$-\sigma\alpha_k\nabla f(x_k)^Td_k+\frac{1}{2}\alpha_k^2d_k^TA_kd_k=$$
$$-\sigma\nabla f(x_k)^Ts_k+\frac{1}{2}s_k^TA_ks_k\to 0. \tag{22}$$

而 $-\sigma\nabla f(x_k)^Ts_k\geqslant 0,\frac{1}{2}s_k^TA_ks_k\geqslant 0$，故有

$-\sigma\nabla f(x_k)^Ts_k\to 0,\frac{1}{2}s_k^TA_ks_k\to 0$. 所以

$$-\sigma\|g_k\|\cdot\|s_k\|\cos\theta_k=\sigma\nabla f(x_k)^Ts_k\to 0. \tag{23}$$

假定 $g_k\to 0$ 不成立，则存在常数 $\varepsilon>0$ 和一个子列使得 $\|g_k\|\geqslant\varepsilon$. 而 $\{d_k\}$ 与 $\{x_k\}$ 是梯度相关的，由性质 2.1 可知，$\cos\theta_k\to 0$ 不成立，则存在 $\varepsilon_0>0$ 和一个子列使得 $|\cos\theta_k|\geqslant\varepsilon_0$. 又 $|\cos\theta_k|\leqslant 1$，由(23)式可知 $\|s_k\|\to 0$，且

$$-g_k^Ts_k=\|g_k\|\cdot\|s_k\|\cdot\cos\theta_k\geqslant\varepsilon\cdot\|s_k\|\cos\theta_k=\varepsilon\cdot\|s_k\|\;|\cos\theta_k|\geqslant\varepsilon\cdot\varepsilon_0\cdot\|s_k\|.$$

但由 Taylor 展式可知，

$$f(x_{k+1})=f(x_k)+g(\xi_k)^Ts_k,$$

其中 ξ_k 位于 x_k 与 x_{k+1} 之间. 由于 $\nabla f(x)$ 在水平集 L 上一致连续，故当 $s_k\to 0$ 时，$g(\xi_k)\to g_k$. 所以

$$f(x_{k+1})=f(x_k)+g_k^Ts_k+o(\|s_k\|),$$

由此可得

$$\frac{f(x_k)-f(x_{k+1})}{-g_k^Ts_k}\to 1+\frac{o(\|s_k\|)}{-g_k^Ts_k}. \tag{24}$$

而

$$0\leqslant\left|\frac{o(\|s_k\|)}{-g_k^Ts_k}\right|\leqslant\frac{o(\|s_k\|)}{\varepsilon\varepsilon_0\|s_k\|},$$

故有

$$0\leqslant\lim_{s_k\to 0}\left|\frac{o(\|s_k\|)}{-g_k^Ts_k}\right|\leqslant\lim_{s_k\to 0}\frac{o(\|s_k\|)}{\varepsilon\varepsilon_0\|s_k\|}=0,$$

所以 $\frac{o(\|s_k\|)}{-g_k^Ts_k}\to 0$. 由(24)式可知

$$\frac{f(x_k)-f(x_{k+1})}{-g_k^Ts_k}\to 1, \tag{25}$$

而

$$0\leqslant\frac{\frac{1}{2}s_k^TA_ks_k}{-g_k^Ts_k}\leqslant\frac{1}{2}\frac{\|s_kA_k\|\cdot\|s_k\|}{\|g_k\|\cdot\|s_k\|\cos\theta_k}\leqslant$$
$$\frac{1}{2}\frac{\|s_kA_k\|\cdot\|s_k\|}{\varepsilon\cdot\varepsilon_0\|s_k\|}=\frac{1}{2}\frac{\|s_kA_k\|}{\varepsilon\cdot\varepsilon_0},$$

所以

$$0\leqslant\lim_{s_k\to 0}\frac{\frac{1}{2}s_k^TA_ks_k}{-g_k^Ts_k}\leqslant\lim_{s_k\to 0}\frac{1}{2}\frac{\|s_kA_k\|}{\varepsilon\cdot\varepsilon_0}=0,$$

故有 $\frac{\frac{1}{2}s_k^TA_ks_k}{-g_k^Ts_k}$. 再由(21)式可知

$$\lim_{s_k\to 0}\frac{f(x_k)-f(x_{k+1})}{-g_k^Ts_k}\leqslant 1-\sigma,$$

这与(25)式矛盾. 故有 $g_k\to 0$.

类似地，当 α_k 采用 WP 准则时定理 3.4 也成立.

定理 3.5 设 $f(x)$ 连续可微，且 $\nabla f(x)$ 满足 Lipischitz 条件

$$\|\nabla f(x)-\nabla f(y)\|\leqslant M\|x-y\|.$$

设在算法中采用 MWP 准则求步长 α_k，如果算法产生的下降方向 $\{d_k\}$ 与点列 $\{x_k\}$ 梯度相关，则要么对某个 k 有 $g_k=0$，要么有 $f(x_k)\to-\infty$，要么有 $g_k\to 0$.

证明 反证法. 假设对每一个 k 都有 $g_k\neq 0$，且 $f(x_k)$ 有下界. 由于 $\{f(x_k)\}$ 单调递减，故 $\{f(x_k)\}$ 的极限存在，因而 $f(x_k)-f(x_{k+1})\to 0$. 由(6)式知

$$f(x_k+\alpha_kd_k)+\frac{1}{2}\alpha_k^2d_k^TA_kd_k-f(x_k)\leqslant$$
$$\sigma\alpha_k\nabla f(x_k)^Td_k,$$

又因为 $f(x_k)-f(x_{k+1})\to 0$，故 $\frac{1}{2}s_k^tA_ks_k-\sigma\cdot s_k^Tg_k\to 0$，其中 $s_k=\alpha_kd_k$. 而 $-\sigma\nabla f(x_k)^Ts_k\geqslant 0,\frac{1}{2}s_k^TA_ks_k\geqslant 0$，故有 $-\sigma g_k^Ts_k\to 0,\frac{1}{2}s_k^TA_ks_k\to 0$. 所以

$$-\sigma\|g_k\|\cdot\|s_k\|\cos\theta_k=\sigma g_k^Ts_k\to 0. \tag{26}$$

假定 $g_k\to 0$ 不成立，则存在常数 $\varepsilon>0$ 和一个子列使得 $\|g_k\|\geqslant\varepsilon$. 因为 $\{d_k\}$ 与 $\{x_k\}$ 是梯度相关的，由性质 2.1 可知，$\cos\theta_k\to 0$ 不成立，则存在 $\varepsilon_0>0$ 和一

个子列使得 $|\cos\theta_k|\geqslant\varepsilon_0$. 又 $|\cos\theta_k|\leqslant 1$,由(26)式可知 $\|s_k\|\to 0$,且

$$-g_k^Ts_k=\|g_k\|\cdot\|s_k\|\cdot\cos\theta_k\geqslant\varepsilon\cdot\|s_k\|\cos\theta_k=\varepsilon\cdot\|s_k\|\,|\cos\theta_k|\geqslant\varepsilon\cdot\varepsilon_0\cdot\|s_k\|.$$

另一方面

$$0<\frac{s_k^TA_kd_k}{\|d_k\|}\leqslant\frac{\|s_k\|\cdot\|A_k\|\cdot\|d_k\|}{\|d_k\|}=\|s_k\|\cdot\|A_k\|,$$

所以

$$0\leqslant\sigma\cdot\|\nabla f(x_k)\|\cdot\cos\theta_k\frac{s_k^TA_kd_k}{M\|d_k\|}\leqslant\frac{\sigma}{M}\|\nabla f(x_k)\|\cdot\cos\theta_k\|s_k\|\cdot\|A_k\|=-\frac{\sigma}{M}\|A_k\|g_k^Ts_k\to 0.$$

由定理 1.2 可知

$$f(x_k)-f(x_k+\alpha_kd_k)\geqslant\frac{\sigma(1-\beta)}{M}\|\nabla f(x_k)\|^2\cos^2\theta_k-\sigma\|\nabla f(x_k)\|\cos\theta_k\cdot\frac{s_k^TA_kd_k}{M\|d_k\|}\geqslant\frac{\sigma(1-\beta)}{M}\varepsilon^2\varepsilon_0^2-\sigma\cdot\|\nabla f(x_k)\|\cos\theta_k\cdot\frac{s_k^TA_kd_k}{M\|d_k\|}.$$

这与 $f(x_k)-f(x_{k+1})\to 0$ 矛盾,从而 $g_k\to 0$.

参考文献:

[1] 袁亚湘,孙文瑜. 最优化理论与方法[M]. 北京:科学出版社,2007.

[2] Dimitri P Bertsekas. Nonlinear programming[M]. USA: Athena Scientific,1999.

[3] Wu Q,Wei Z. Some new step-size rules for optimization problems[J]. Journal of Shanghai University(English Edition),2007,11(2):135-141.

[4] Qi L. Convergence analysis of some algorithms for solving nonsmooth equations[J]. Mathematics of Operationd Research,1993,18(1):227-244.

[5] 韦增欣,武小平,刘利英. 广义 Wolfe 线搜索下共轭梯度法的全局收敛性[J]. 广西科学,2005,12(3):167-169.

[6] Wei Z,Qi L,Jiang H. Some convergence properties of descent methods[J]. J Optim Theory Appl,1997,95(1):177-188.

[7] 陆春桃. 一些修正的线搜索及其收敛性[J]. 广西师范学院学报:自然科学版,2006,6(2):13-18.

[8] 韦增欣. 一类特殊算法的收敛性质[J]. 曲阜师范大学学报,1991,17(2):10-14.

[9] Mokhtar S Bazaraa,Hanif D Sherali,C M Shetty. Nonlinear programming theory and algorithms(second edition)[M]. Canada:John Wiley&Sons,Inc ,1993.

(责任编辑:尹 闯)

科学家发现影响生物钟节律蛋白质

很多植物春季开花,秋季结果;夜行动物白天睡大觉,夜晚则四处“狩猎”。例如,十字花科植物——拟南芥的生物周期约为 24h;果蝇通常白天活动,夜晚休息。决定这些生理节律的生物周期被称为“生物钟”。最近,科学家培育出一种生物周期达到 72h 的拟南芥,并且还通过红外线照射使果蝇在白天和夜晚都四处活动。科学家重点检查与这些拟南芥和果蝇的生物钟有关的生命活动物质,发现它们的 PRMT5 蛋白质都发生了变异。事实上,PRMT5 蛋白质通过调控某些生物的基因转录、核糖核酸剪切和细胞增殖,保证有关生物正常生长发育;该蛋白质变异可以导致与开花等重要生命活动有关的植物基因表达发生改变,使得开花提前或推迟,影响其正常生物节律;同样,与 PRMT5 蛋白质变异有关的另外一些基因变化,也会引起动物与时间概念有关行为的改变。这些发现有具体应用价值,尤其是在农业方面。比如某些植物的叶子越多,收成越好,而叶子数量取决于花期长短,花期则由生物钟控制。如果通过基因调控影响这些植物的生物钟,就有望使它们长出更多的叶子,带来更好的经济效益。

(据科学网)