不精确线搜索

寇彩霞

Email: koucx@bupt.edu.cn

北京邮电大学理学院 主楼-816

精确线搜索

基本假设：

1. $f(x)$ 在$\mathcal{R}^n$上连续可微；
2. 梯度$g(x)$满足Lip条件

$$\|g(x)-g(y)\|_2 \leq \|x-y\|_2.$$

- α_k的下界
- 函数的下降量
- 收敛性

基本假设：

1. $f(x)$ 在$\mathcal{R}^n$上连续可微；
2. 梯度$g(x)$满足Lip条件

$$\|g(x) - g(y)\|_2 \leq \|x - y\|_2.$$

- α_k的下界
- 函数的下降量
- 收敛性

基本假设：

1. $f(x)$ 在$\mathcal{R}^n$上连续可微；
2. 梯度$g(x)$满足Lip条件

$$\|g(x) - g(y)\|_2 \leq \|x - y\|_2.$$

- α_k的下界
- 函数的下降量
- 收敛性

背景

- 精确线搜索方法求
$$\varphi(\alpha) = f(x_k + \alpha d_k)$$
的精确极小点，花费的计算量较大. 一般在迭代过程中，没有必要把线性搜索搞得十分精确.
- 当迭代点离目标函数的最优解尚远时，过分追求线性搜索的精度反而会降低整个算法的效率. 另外，一些最优化方法，其收敛速度并不依赖于精确的一维搜索过程.
- 放松对α_k的精确度要求，只要求目标函数在迭代的每一步都有充分的下降即可，这样可以大大节省工作量.
- 基本思想：要求满足某种“充分下降”条件.

- Goldstein准则为

$$f(x_k+\alpha_k d_k) \le f(x_k)+\rho\alpha_k g_k^T d_k, \tag{3.5.1}$$

$$f(x_k+\alpha_k d_k) \ge f(x_k)+(1-\rho)\alpha_k g_k^T d_k, \tag{3.5.2}$$

其中, $0<\rho<\frac{1}{2}$.

- 第一个不等式是充分下降条件，第二个不等式保证了α_k不会取得太小，因为当α_k取得太小时，算法前进很慢.
- 若设$\varphi(\alpha)=f(x_k+\alpha d_k)$, 则(3.5.1)和(3.5.2)可以分别写成

$$\varphi(\alpha_k) \le \varphi(0)+\rho\alpha_k\varphi'(0), \tag{3.5.3}$$

$$\varphi(\alpha_k) \ge \varphi(0)+(1-\rho)\alpha_k\varphi'(0). \tag{3.5.4}$$

- Goldstein准则为

$$f(x_k + \alpha_k d_k) \le f(x_k) + \rho \alpha_k g_k^T d_k, \qquad (3.5.1)$$

$$f(x_k + \alpha_k d_k) \ge f(x_k) + (1 - \rho) \alpha_k g_k^T d_k, \qquad (3.5.2)$$

其中, $0 < \rho < \frac{1}{2}$.

- 第一个不等式是充分下降条件，第二个不等式保证了α_k不会取得太小，因为当α_k取得太小时，算法前进很慢.
- 若设$\varphi(\alpha) = f(x_k + \alpha d_k)$, 则(3.5.1)和(3.5.2)可以分别写成

$$\varphi(\alpha_k) \le \varphi(0) + \rho \alpha_k \varphi'(0), \qquad (3.5.3)$$

$$\varphi(\alpha_k) \ge \varphi(0) + (1 - \rho) \alpha_k \varphi'(0). \qquad (3.5.4)$$

- Goldstein准则为

$$f(x_k+\alpha_k d_k) \le f(x_k)+\rho\alpha_k g_k^T d_k, \tag{3.5.1}$$

$$f(x_k+\alpha_k d_k) \ge f(x_k)+(1-\rho)\alpha_k g_k^T d_k, \tag{3.5.2}$$

其中, $0<\rho<\frac{1}{2}$.

- 第一个不等式是充分下降条件，第二个不等式保证了α_k不会取得太小，因为当α_k取得太小时，算法前进很慢.
- 若设$\varphi(\alpha)=f(x_k+\alpha d_k)$, 则(3.5.1)和(3.5.2)可以分别写成

$$\varphi(\alpha_k) \le \varphi(0)+\rho\alpha_k\varphi'(0), \tag{3.5.3}$$

$$\varphi(\alpha_k) \ge \varphi(0)+(1-\rho)\alpha_k\varphi'(0). \tag{3.5.4}$$

- 算法3.5.1
  - 步1. 选取初始数据. 给出初始搜索区间 $[a_0, b_0]$,($[0, +\infty)$ 或 $[0, \alpha_{\max}]$), 取定初始点$\alpha_0 \in [a_0, b_0]$. 计算$\varphi(0), \varphi'(0)$, 给出$\rho \in \left(0, \frac{1}{2}\right)$, $t > 1, k := 0$.
  - 步2. 检验准则(3.5.3). 计算$\varphi(\alpha_k)$. 若

$$\varphi(\alpha_k) \leq \varphi(0) + \rho\alpha_k\varphi'(0),$$

转步3；否则，令$a_{k+1} := a_k, b_{k+1} := \alpha_k$, 转步4.
  - 步3. 检验准则(3.5.4). 若

$$\varphi(\alpha_k) \geq \varphi(0) + (1-\rho)\alpha_k\varphi'(0),$$

停止迭代，输出α_k; 否则，令$a_{k+1} := \alpha_k, b_{k+1} := b_k$.
若$b_{k+1} < +\infty$(或$\alpha_{\max}$)转步4; 否则，令$\alpha_{k+1} := t\alpha_k, k := k+1$, 转步2.
  - 步4. $\alpha_{k+1} := \dfrac{a_{k+1} + b_{k+1}}{2}, k := k+1$, 转步2.

- 用Goldstein法求解

$$\min_{t\geq 0} \varphi(t) = t^3 - 2t + 1$$

取$\alpha_0 = 2, \rho = 0.2, t = 2.$

- 在Goldstein准则中，(3.5.2)的一个缺点是可能把$\varphi(\alpha)=f(x_k+\alpha d_k)$的极小点排除在可接受的区间以外.
- 为了克服这一缺点并同时保证α_k不是太小，Wolfe提出了下面的条件代替(3.5.2)：

$$g_{k+1}^T d_k \geq \sigma g_k^T d_k, \sigma \in (\rho, 1), \tag{3.5.5}$$

即

$$\begin{aligned}\varphi'(\alpha_k) &= g(x_k+\alpha_k d_k)^T d_k \geq \sigma g_k^T d_k \\ &= \sigma\varphi'(0) > \varphi'(0),\end{aligned} \tag{3.5.6}$$

其几何解释是在可接受点处切线的斜率$\varphi'(\alpha_k)$大于或等于初始斜率的σ倍. 这个条件也叫做曲率条件.

WOLFE准则

- 在Goldstein准则中，(3.5.2)的一个缺点是可能把$\varphi(\alpha) = f(x_k + \alpha d_k)$的极小点排除在可接受的区间以外.
- 为了克服这一缺点并同时保证α_k不是太小，Wolfe提出了下面的条件代替(3.5.2)：

$$g_{k+1}^T d_k \geq \sigma g_k^T d_k, \sigma \in (\rho, 1), \tag{3.5.5}$$

即

$$\begin{aligned} \varphi'(\alpha_k) &= g(x_k + \alpha_k d_k)^T d_k \geq \sigma g_k^T d_k \\ &= \sigma \varphi'(0) > \varphi'(0), \end{aligned} \tag{3.5.6}$$

其几何解释是在可接受点处切线的斜率$\varphi'(\alpha_k)$大于或等于初始斜率的σ倍. 这个条件也叫做曲率条件.

- 这样，充分下降条件和曲率条件一起构成了Wolfe准则：

$$f(x_k + \alpha_k d_k) \le f(x_k) + \rho \alpha_k g_k^T d_k, \tag{3.5.7}$$

$$g(x_k + \alpha_k d_k)^T d_k \ge \sigma g_k^T d_k, \tag{3.5.8}$$

  其中，$0 < \rho < \sigma < 1$.
- 不等式(3.5.8)是精确线搜索所满足的正交条件

$$g_{k+1}^T d_k = 0$$

  的近似. 但(3.5.5)的不足之处是即使在$\sigma \to 0$时也不能导致精确线性搜索.

- 针对Wolfe条件的不足，提出强Wolfe准则：

$$f(x_k + \alpha_k d_k) \le f(x_k) + \rho\alpha_k g_k^T d_k, \qquad (3.5.9)$$

$$|g(x_k + \alpha_k d_k)^T d_k| \le \sigma|g_k^T d_k|, \qquad (3.5.10)$$

其中，$0 < \rho < \sigma < 1$. 这样，当$\sigma \to 0$时，(3.5.10)的极限便是精确线性搜索.

- 一般σ越小，线性搜索越精确，但同时工作量越大. 而不精确线性搜索不要求过小的σ, 通常取$\rho = 0.1, \sigma \in [0.6, 0.8]$.

- 针对Wolfe条件的不足，提出强Wolfe准则：

$$f(x_k + \alpha_k d_k) \le f(x_k) + \rho \alpha_k g_k^T d_k, \tag{3.5.9}$$

$$|g(x_k + \alpha_k d_k)^T d_k| \le \sigma |g_k^T d_k|, \tag{3.5.10}$$

其中，$0 < \rho < \sigma < 1$. 这样，当$\sigma \to 0$时，(3.5.10)的极限便是精确线性搜索.

- 一般σ越小，线性搜索越精确，但同时工作量越大. 而不精确线性搜索不要求过小的σ, 通常取$\rho = 0.1, \sigma \in [0.6, 0.8]$.

- 步1. 选取初始数据. 给定初始搜索区间$[0, \alpha_{\max}]$, $\rho \in (0, \frac{1}{2})$, $\sigma \in (\rho, 1)$.
  令$\alpha_1 = 0, \alpha_2 = \alpha_{\max}$, 计算$\varphi_1 = f(x_k), \varphi_1' = g(x_k)^T d_k$, 取$\alpha \in (0, \alpha_2)$.
- 步2. 计算$\varphi = \varphi(\alpha) = f(x_k + \alpha d_k)$. 若$f(x_k + \alpha d_k) \le f(x_k) + \rho \alpha g_k^T d_k$, 转步3；否则，由二次插值公式(3.3.14)计算$\overline{\alpha}$:

$$\overline{\alpha} = \alpha_1 + \frac{\alpha - \alpha_1}{2\left(1 + \dfrac{\varphi_1 - \varphi}{(\alpha - \alpha_1)\varphi_1'}\right)}.$$

  令$\alpha_2 := \alpha, \alpha := \overline{\alpha}$. 转步2.
- 步3. 计算$\varphi' = \varphi'(\alpha) = g(x_k + \alpha d_k)^T d_k$. 若$g(x_k + \alpha d_k)^T d_k \ge \sigma g_k^T d_k$, 则令$\alpha_k = \alpha$, 输出$\alpha_k$, 停；否则，由二次插值公式(3.3.17)计算$\overline{\alpha}$:

$$\overline{\alpha} = \alpha + \frac{(\alpha - \alpha_1)\varphi'}{\varphi_1' - \varphi'}.$$

  令$\alpha_1 := \alpha, \varphi_1 := \varphi, \varphi_1' := \varphi', \alpha := \overline{\alpha}$, 转步2.

- 设d_k是$f(x)$在x_k处的下降方向，给定$\beta \in (0,1), \rho \in (0,\frac{1}{2}), \tau > 0$. 设$m_k$是使得下述不等式

$$f(x_k + \beta^m \tau d_k) \le f(x_k) + \rho \beta^m \tau g_k^T d_k. \tag{3.5.11}$$

成立的最小非负整数. 由于d_k是下降方向，当m充分大时，不等式(3.5.11)总是成立的, 因此上述m_k总存在.

- 令

$$\alpha_k = \beta^{m_k} \tau. \tag{3.5.12}$$

由于m_k是使得不等式(3.5.11)成立的最小非负整数，因而α_k不会太小. 上式也保证了目标函数$f(x)$的充分下降. 实际上(3.5.11)就是充分下降条件

$$\varphi(\alpha_k) \le \varphi(0) + \rho \alpha_k \varphi'(0). \tag{3.5.13}$$

- 设d_k是$f(x)$在x_k处的下降方向，给定$\beta \in (0,1), \rho \in (0,\frac{1}{2}), \tau > 0$. 设$m_k$是使得下述不等式

$$f(x_k + \beta^m \tau d_k) \le f(x_k) + \rho \beta^m \tau g_k^T d_k. \tag{3.5.11}$$

成立的最小非负整数. 由于d_k是下降方向，当m充分大时，不等式(3.5.11)总是成立的, 因此上述m_k总存在.

- 令

$$\alpha_k = \beta^{m_k} \tau. \tag{3.5.12}$$

由于m_k是使得不等式(3.5.11)成立的最小非负整数，因而α_k不会太小. 上式也保证了目标函数$f(x)$的充分下降. 实际上(3.5.11)就是充分下降条件

$$\varphi(\alpha_k) \le \varphi(0) + \rho \alpha_k \varphi'(0). \tag{3.5.13}$$

- 如果

$$\varphi(\alpha_k) \leq \varphi(0) + \rho\alpha_k\varphi'(0), \tag{3.5.13}$$

满足，则终止搜索；否则，缩小α_k, 或者在区间$[0, \alpha_k]$上用二次插值公式(3.3.15)求近似极小点$\overline{\alpha}$

$$\overline{\alpha} = -\frac{1}{2}\frac{\varphi'(0)\alpha_k^2}{\varphi(\alpha_k) - \varphi(0) - \varphi'(0)\alpha_k}. \tag{3.5.14}$$

将其作为新的α_k, 这是一个插值法与充分下降条件组合起来的线性搜索方法.

- 开始时, 令 $\alpha = 1$, 如果 $x_k + \alpha d_k$ 不可接受, 则减少 α (即后退), 一直到 $x_k + \alpha d_k$ 可接受为止. 该方法也叫后退方法(Back- ward).

- 算法 3.5.3
- 步0. 给出 $\rho \in (0, \frac{1}{2})$, $0 < l < u < 1$.
- 步1. 取 $\alpha = 1$.
- 步2. 检验

$$f(x_k + \alpha d_k) \leq f(x_k) + \rho \alpha g_k^T d_k \tag{3.5.15}$$

是否满足.

- 步3. 如果 (3.5.15) 不满足, 取 $\alpha := \omega\alpha$, 其中 $\omega \in [l, u]$, 转步2. 否则, 取 $\alpha_k = \alpha$, $x_{k+1} := x_k + \alpha_k d_k$.

- 为保证方法的下降性，我们要求避免搜索方向$s_k = \alpha_k d_k$和负梯度方向$-g_k$几乎正交的情形. 否则$s_k^T g_k$接近于零，s_k几乎不是下降方向.
- 假定s_k和$-g_k$的夹角θ_k满足

$$\theta_k \leq \frac{\pi}{2} - \mu, \quad \forall k, \tag{3.6.1}$$

其中，$\mu > 0, \theta_k \in [0, \frac{\pi}{2}]$定义为

$$\cos \theta_k = \frac{-g_k^T s_k}{\|g_k\| \|s_k\|}. \tag{3.6.2}$$

采用不精确线性搜索准则的一般下降算法

- 算法 3.6.1
- 步1. 给出 $x_0 \in R^n$, $0 \le \varepsilon < 1$, $k := 0$.
- 步2. 如果 $\|g_k\| \le \varepsilon$, 停止. 否则, 求出下降方向 d_k, 使其满足 $d_k^T g_k < 0$.
- 步3. 利用 Goldstein 准则 (3.5.1) – (3.5.2) 或 Wolfe 准则 (3.5.7) – (3.5.8) 求出步长因子 α_k.
- 步4. 令 $x_{k+1} = x_k + \alpha_k d_k$; $k := k + 1$, 转步2.

以Wolfe搜索为例，给出不精确线性搜索在单步中至少下降的界.

引理 (3.6.2)

设函数 $f(x)$ 连续可微，梯度 $g(x)$ 满足 *Lipschitz* 连续条件

$$\|g(y) - g(z)\| \leq M\|y - z\|, \tag{3.6.3}$$

如果 $f(x_k + \alpha d_k)$ 下有界，$\alpha > 0$, 则对满足 *Wolfe* 准则 (3.5.7) – (3.5.8) 的任何 $\alpha_k > 0$ 均有

$$f(x_k) - f(x_k + \alpha_k d_k) \geq \frac{\rho(1-\sigma)}{M}\|g_k\|^2 \cos^2\langle d_k, -g_k\rangle. \tag{3.6.4}$$

采用GOLDSTEIN准则的一般下降算法的总体收敛性

定理 (3.6.3)

设在算法 *3.6.1* 中采用 *Goldstein* 准则 (3.5.1) – (3.5.2) 求步长因子 α_k, 并设夹角条件 (3.6.1) 满足. 如果 $g(x)$ 存在, 且在水平集 $\{x| \ f(x) \le f(x_0)\}$ 上一致连续, 那么, 或者对某个 k, 有 $g_k = 0$, 或者 $f(x_k) \to -\infty$, 或者 $g_k \to 0$.

定理 (3.6.4)

设 $f : R^n \to R$ 在 R^n 上连续可微和下有界. $g(x)$ 在水平集 $\Omega = \{x | f(x) \le f(x_0)\}$ 上一致连续. 设不精确线性搜索方法采用 *Wolfe* 准则 (3.5.7) – (3.5.8), 则

$$\lim_{k\to\infty} \|g_k\| \cos\theta_k = 0. \tag{3.6.6}$$

如果夹角条件 (3.6.1) 满足，则

$$\lim_{k\to\infty} g_k = 0. \tag{3.6.7}$$

小结

- 不精确线性搜索方法

  Goldstein准则， Wolfe准则， Armijo准则
- 不精确线性搜索的收敛性