3-2 非精确线性搜索

非精确线性搜索

精确线性搜索存在问题：

- 当维数n比较大且f比较复杂时，每步迭代精确获得步长涉及计算量很大。
- 若当前迭代点离最优解尚远时，没必要做高精度线性搜索。
- 针对某些非光滑函数或导数表达式复杂的函数不适用。

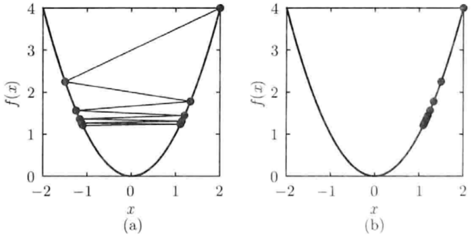
线性搜索方法每步确定**合适的步长**α **(非精确线性搜索)**，保证满意的**函数值下降**。因而大大减少了计算量，提高整体迭代算法效率。代表性非精确线性搜索准则有：

- Armijo准则
- Goldstein准则
- Wolfe准则

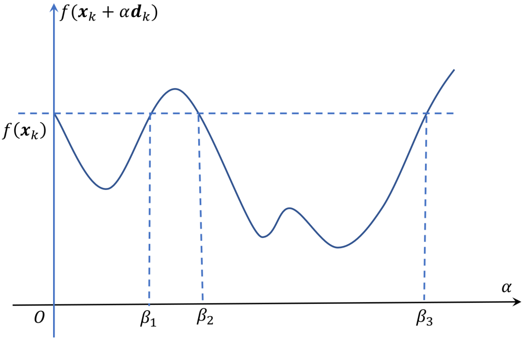
例3.2.1 利用表格中搜索方向以及步长求解问题：$\min f(x) = x^2$。

	方法1	方法2
步长α_k	$2 + \frac{2k+1}{k(k+1)}$	$\frac{1}{k(k+1)}$
方向d_k	$(-1)^k$	-1
迭代点x_k	$(-1)^{k+1}(1 + \frac{1}{k})$	$1 + \frac{1}{k}$
聚点	$-1, 1$	1

解： 从$x_1 = 2$出发, 两种方法产生的点列函数值下降，但不收敛。



究竟选择什么样的步长合适？讨论一般线性搜索问题$f(\boldsymbol{x}_k+\alpha\boldsymbol{d}_k)$．满足下降的步长区间为：$(0,\beta_1)\cup(\beta_2,\beta_3)$。方法1使用的步长太接近于区间$(\beta_2,\beta_3)$的上界，而方法2选用的步长太接近区间$(0,\beta_1)$的下界。下降不明显或者移动不充分导致不收敛。

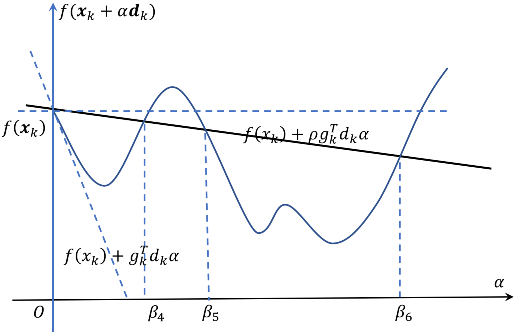


选用的步长不应太接近0或者太接近β_3。

Armijo准则

设$\boldsymbol{d}_k$是$f(\boldsymbol{x})$在$\boldsymbol{x}_k$处的下降方向($\boldsymbol{g}_k^T\boldsymbol{d}_k < 0$), $\rho \in (0,1)$, 步长α应满足如下不等式:

$$f(\boldsymbol{x}_k + \alpha\boldsymbol{d}_k) \leq f(\boldsymbol{x}_k) + \rho\boldsymbol{g}_k^T\boldsymbol{d}_k\alpha. \tag{3.2.1}$$



建议$\rho = 10^{-3}$或更小的值, 满足下降的α取值范围$(0,\beta_4],[\beta_5,\beta_6]$。

Armijo准则

如何确定α?

设：

$$\alpha = \beta^{m_k}\tau.$$

其中: $\beta \in (0,1)$, $\tau > 0$。m_k是使得(3.2.1)满足的最小非负整数。

由于$\boldsymbol{d}_k$是下降方向,

- 当m_k充分大时，不等式(3.2.1)总是成立的(保证下降)，故上述m_k总是存在的。
- 由于m_k是使得下降不等式成立的最小非负整数，因而α不会太小，从而保证了目标函数$f(\boldsymbol{x})$的充分下降。

注意： 不能保证步长不接近于零。

Goldstein准则

为避免选取过小的步长，Goldstein准则选取步长α应满足以下两个条件：

$$f(\boldsymbol{x}_k+\alpha\boldsymbol{d}_k)\leq f(\boldsymbol{x}_k)+\rho\alpha\boldsymbol{g}_k^T\boldsymbol{d}_k, \tag{3.2.2}$$

$$f(\boldsymbol{x}_k+\alpha\boldsymbol{d}_k)\geq f(\boldsymbol{x}_k)+(1-\rho)\alpha\boldsymbol{g}_k^T\boldsymbol{d}_k, \tag{3.2.3}$$

其中，$0<\rho<\frac{1}{2}$.

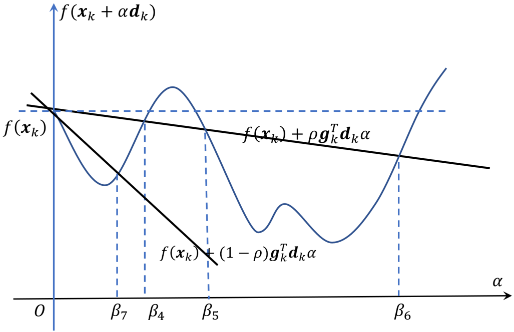
- 第一个不等式保证充分下降条件;
- 第二个不等式保证了α不会取得太小。

两个条件(3.2.2)和(3.2.3)可以用$\varphi(\alpha)$表示为:

$$\varphi(\alpha)\leq\varphi(0)+\rho\alpha\varphi'(0), \tag{3.2.4}$$

$$\varphi(\alpha)\geq\varphi(0)+(1-\rho)\alpha\varphi'(0). \tag{3.2.5}$$

Goldstein准则



满足条件的步长区间$[\beta_7, \beta_4]$和$[\beta_5, \beta_6]$。

Goldstein非精确线性搜索算法步骤

算法3.2.1 – Goldstein非精确线性搜索算法

步1 选取初始数据。在初始搜索区间$[0,+\infty)$（或$[0,\alpha_{max}]$）中取初始点α_0, 计算$\varphi(0)$, $\varphi'(0)$, 给出$\rho\in(0,\frac{1}{2})$, $t>1$, $k:=0$.

步2 检验准则(3.2.4).计算$\varphi(\alpha_k)$. 若

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

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(3.2.5). 若

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

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 $\alpha_{k+1}:=\frac{a_{k+1}+b_{k+1}}{2}$, $k:=k+1$, 转步2.

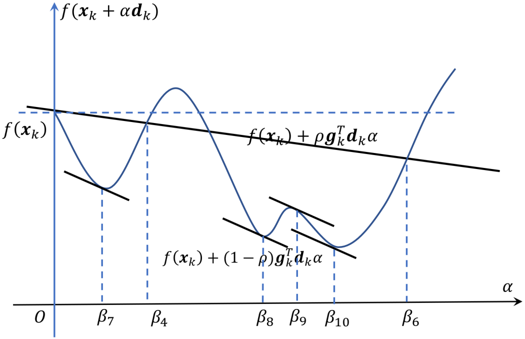
Wolfe 准则

Goldstein准则一个缺点是可能把$\varphi(\alpha)$的极小点排除在可接受区间之外。Wolfe准则需满足以下下降条件和曲率条件:

$$f(\boldsymbol{x}_k + \alpha \boldsymbol{d}_k) \le f(\boldsymbol{x}_k) + \rho\alpha \boldsymbol{g}_k^T \boldsymbol{d}_k, \tag{3.2.6}$$

$$g(\boldsymbol{x}_k + \alpha \boldsymbol{d}_k)^T \boldsymbol{d}_k \ge \sigma \boldsymbol{g}_k^T \boldsymbol{d}_k, \tag{3.2.7}$$

其中, $0 < \rho < \sigma < 1$.



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

应该指出，曲率条件可视为精确线性搜索所满足的正交条件$\boldsymbol{g}_{k+1}^T\boldsymbol{d}_k=0$的近似，但$\sigma\to 0$时并不能导致精确线性搜索。

强Wolfe准则：

$$f(\boldsymbol{x}_k+\alpha\boldsymbol{d}_k)\le f(\boldsymbol{x}_k)+\rho\alpha\boldsymbol{g}_k^T\boldsymbol{d}_k, \tag{3.2.8}$$

$$|\boldsymbol{g}(\boldsymbol{x}_k+\alpha_k\boldsymbol{d}_k)^T\boldsymbol{d}_k|\le\sigma|\boldsymbol{g}_k^T\boldsymbol{d}_k|, \tag{3.2.9}$$

其中, $0<\rho<\sigma<1.$

当$\sigma\to 0$时，(3.2.9)的极限便是精确线性搜索。一般地, σ愈小, 线性搜索愈精确. 不过, σ愈小,工作量愈大。非精确线性搜索σ取值不能太小，通常取$\rho=0.1$, $\sigma\in[0.6,0.8]$。

非精确线性搜索下降算法步骤

算法3.2.2–非精确线性搜索下降算法

步1 给出$\boldsymbol{x}_0 \in \mathbb{R}^n$, $0 \le \varepsilon < 1$, $k := 0$.

步2 如果$\|\boldsymbol{g}_k\| \le \varepsilon$,则停止；否则，确定下降方向$\boldsymbol{d}_k$，使其满足：

$$\boldsymbol{d}_k^T \boldsymbol{g}_k < 0.$$

步3 利 用Goldstein准 则(3.2.4)-(3.2.5)或Wolfe准 则(3.2.6)-(3.2.7)求 出 步长因子α_k.

步4 令$\boldsymbol{x}_{k+1} = \boldsymbol{x}_k + \alpha_k \boldsymbol{d}_k$; $k := k+1$,转步2.

注意： 搜索方向$\boldsymbol{d}_k$满足$\boldsymbol{d}_k^T \boldsymbol{g}_k < 0$保证$\boldsymbol{d}_k$为下降方向。

类似于**定理**3.2.1, 以Wolfe准则为例，给出非精确线性搜索在单步中函数值下降的界。

定理3.2.1 单步中函数值下降界估计.

设函数$f(\boldsymbol{x})$连续可微，梯度$\boldsymbol{g}(\boldsymbol{x})$满足Lipschitz连续条件:

$$\|g(\boldsymbol{y}) - g(\boldsymbol{z})\| \leq M\|\boldsymbol{y} - \boldsymbol{z}\|$$

如果$f(\boldsymbol{x}_k + \alpha \boldsymbol{d}_k)$下有界，$\alpha > 0$,则对满足Wolfe准则(3.2.6)-(3.2.7)的任何$\alpha_k > 0$均有:

$$f(\boldsymbol{x}_k) - f(\boldsymbol{x}_k + \alpha_k \boldsymbol{d}_k) \geq \frac{\rho(1-\sigma)}{M}\|\boldsymbol{g}_k\|^2 \cos^2\langle \boldsymbol{d}_k, -\boldsymbol{g}_k\rangle.$$

定理3.2.1证明

证明： 由Lipschitz条件和Wolfe准则曲率条件(3.2.7), 得：

$$\alpha_k M\|\boldsymbol{d}_k\|^2 \geq \boldsymbol{d}_k^T(\boldsymbol{g}_{k+1}-\boldsymbol{g}_k) \geq -(1-\sigma)\boldsymbol{d}_k^T\boldsymbol{g}_k$$

进一步有:

$$\alpha_k\|\boldsymbol{d}_k\| \geq \frac{1-\sigma}{M\|\boldsymbol{d}_k\|}\|\boldsymbol{d}_k\|\|\boldsymbol{g}_k\|\cos\langle\boldsymbol{d}_k,-\boldsymbol{g}_k\rangle = \frac{1-\sigma}{M}\|\boldsymbol{g}_k\|\cos\langle\boldsymbol{d}_k,-\boldsymbol{g}_k\rangle$$

结合Wolfe准则下降条件(3.2.6),有：

$$\begin{aligned} f(\boldsymbol{x}_k)-f(\boldsymbol{x}_k+\alpha_k\boldsymbol{d}_k) &\geq -\alpha_k\rho\boldsymbol{d}_k^T\boldsymbol{g}_k = \alpha_k\rho\|\boldsymbol{d}_k\|\|\boldsymbol{g}_k\|\cos\langle\boldsymbol{d}_k,-\boldsymbol{g}_k\rangle \\ &\geq \rho\|\boldsymbol{g}_k\|\cos\langle\boldsymbol{d}_k,-\boldsymbol{g}_k\rangle\frac{1-\sigma}{M}\|\boldsymbol{g}_k\|\cos\langle\boldsymbol{d}_k,-\boldsymbol{g}_k\rangle \\ &= \frac{\rho(1-\sigma)}{M}\|\boldsymbol{g}_k\|^2\cos^2\langle\boldsymbol{d}_k,-\boldsymbol{g}_k\rangle \end{aligned}$$

结论得证。 ■

非精确线性搜索总体收敛

定理3.2.2 Goldstein准则总体收敛性.

设在**算法**3.1.1中采用Goldstein准则(3.2.4)–(3.2.5)确定步长因子α_k，且满足夹角条件:

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

如果$\boldsymbol{g}(\boldsymbol{x})$存在,且在水平集$L(\boldsymbol{x}_0) = \{\boldsymbol{x} | f(\boldsymbol{x}) \leq f(\boldsymbol{x}_0)\}$上一致连续，则以下结论之一成立：(1) 对某个$N$,有$\boldsymbol{g}_N = 0$; (2) $f(\boldsymbol{x}_k) \to -\infty$; (3) $\boldsymbol{g}_k \to 0$。

这里：

$$\mu > 0, \theta_k \in [0, \frac{\pi}{2}], \cos\theta_k = \frac{-\boldsymbol{g}_k^T \boldsymbol{s}_k}{\|\boldsymbol{g}_k\| \|\boldsymbol{s}_k\|}$$

定理3.2.2证明

证明： 假定$f(\boldsymbol{x}_k)$下有界，且对所有的k，$\boldsymbol{g}_k=\nabla f(\boldsymbol{x}_k)\neq 0$，知下降方向非零$\boldsymbol{s}_k\neq 0$和序列$\{f(\boldsymbol{x}_k)\}$单调下降，故$\{f(\boldsymbol{x}_k)\}$有极限，进而有

$$f(\boldsymbol{x}_k)-f(\boldsymbol{x}_{k+1})\to 0$$

由Goldstein下降条件(3.2.4)得到$-\boldsymbol{g}_k^T\boldsymbol{s}_k\to 0.$

假定$\boldsymbol{g}_k\to 0$不成立，则存在$\varepsilon>0$和一个子列使得$\|\boldsymbol{g}_k\|\geq\varepsilon$ 和$\|\boldsymbol{s}_k\|\to 0.$
由于$\theta_k\leq\frac{\pi}{2}-\mu$,故

$$\cos\theta_k\geq\cos(\frac{\pi}{2}-\mu)=\sin\mu,$$

因而

$$-\boldsymbol{g}_k^T\boldsymbol{s}_k\geq\sin\mu\|\boldsymbol{g}_k\|\|\boldsymbol{s}_k\|\geq\varepsilon\sin\mu\|s_k\|.$$

定理3.2.2证明

证明续： 由Taylor公式给出：

$$f(\boldsymbol{x}_{k+1}) = f(\boldsymbol{x}_k) + \boldsymbol{g}(\boldsymbol{\xi}_k)^T \boldsymbol{s}_k,$$

其中$\boldsymbol{\xi}_k$位于线段$(\boldsymbol{x}_k, \boldsymbol{x}_{k+1})$上. 由$\boldsymbol{g}(\boldsymbol{x})$的一致连续性,当$\boldsymbol{s}_k \to 0$时,$\boldsymbol{g}(\boldsymbol{\xi}_k) \to \boldsymbol{g}_k$。
此时,

$$f(\boldsymbol{x}_{k+1}) = f(\boldsymbol{x}_k) + \boldsymbol{g}_k^T \boldsymbol{s}_k + o(\|\boldsymbol{s}_k\|).$$

由此有

$$\frac{f(\boldsymbol{x}_k) - f(\boldsymbol{x}_{k+1})}{-\boldsymbol{g}_k^T \boldsymbol{s}_k} \to 1,$$

这与Goldstein准则(3.2.5)条件矛盾. 故$\boldsymbol{g}_k \to 0$. 结论(3)得证。 ■

Wolfe准则整体收敛性

定理3.2.3 Wolfe准则整体收敛性.

设$f:\mathbb{R}^n \to \mathbb{R}$在$\mathbb{R}^n$上连续可微，在水平集$L(\boldsymbol{x}_0)=\{\boldsymbol{x}|f(\boldsymbol{x})\le f(\boldsymbol{x}_0)\}$，$f$有下界，$\boldsymbol{g}(\boldsymbol{x})=\nabla f(\boldsymbol{x})$一致连续. 设**算法**3.1.2采用Wolfe准则(3.2.6)–(3.2.7),则

$$\lim_{k\to\infty}\|\boldsymbol{g}_k\|\cos\theta_k=0. \tag{3.2.10}$$

如果**定理**3.2.2中夹角条件满足,则

$$\boldsymbol{g}_k\to 0, k\to\infty \tag{3.2.11}$$

证明： 由于$\boldsymbol{g}_k^T\boldsymbol{s}_k<0$，又由于$f$下有界，因此序列$\{\boldsymbol{x}_k\}$是有定义的，且在水平集$L(\boldsymbol{x}_0)$中. **反证法** 假定(3.2.10)不成立，则存在$\varepsilon>0$和子序列，其指标集为$\mathcal{K}$,使得:

$$\|\boldsymbol{g}_k\|\cos\theta_k=\|\boldsymbol{g}_k\|\frac{-\boldsymbol{g}_k^T\boldsymbol{s}_k}{\|-\boldsymbol{g}_k\|\|\boldsymbol{s}_k\|}=-\frac{\boldsymbol{g}_k^T\boldsymbol{s}_k}{\|\boldsymbol{s}_k\|}\ge\varepsilon,\forall k\in\mathcal{K}.$$

定理3.2.3证明

证明续： 由Wolfe准则下降条件(3.2.6)得：

$$f(\boldsymbol{x}_k) - f(\boldsymbol{x}_{k+1}) \geq \rho\|\boldsymbol{s}_k\| \left(-\frac{\boldsymbol{g}_k^T \boldsymbol{s}_k}{\|\boldsymbol{s}_k\|}\right) \geq \rho\|\boldsymbol{s}_k\|\varepsilon.$$

又由于$\{f(\boldsymbol{x}_k)\}$单调下降,因而是收敛的，故$\{\boldsymbol{s}_k\}_{\mathcal{K}}$收敛到零.
又由Wolfe准则曲率条件(3.2.7),

$$(1-\sigma)(-\boldsymbol{g}_k^T \boldsymbol{s}_k) \leq (\boldsymbol{g}_{k+1} - \boldsymbol{g}_k)^T \boldsymbol{s}_k, k \geq 0.$$

因此,

$$\varepsilon \leq -\frac{\boldsymbol{g}_k^T \boldsymbol{s}_k}{\|\boldsymbol{s}_k\|} \leq \frac{1}{1-\sigma}\|\boldsymbol{g}_{k+1} - \boldsymbol{g}_k\|, k \in \mathcal{K}.$$

由于$\{\boldsymbol{s}_k\}_{\mathcal{K}}$收敛到零，故由$\boldsymbol{g}(\boldsymbol{x})$在水平集上一致连续知上式右边趋于零，从而产生矛盾.这样,(3.2.10)得证.

定理3.2.3证明

证明续： 进一步,若夹角条件满足, 则存在一个正数δ使得

$$\cos\theta_k \geq \delta > 0, \quad \forall k. \tag{3.2.12}$$

(3.2.10)和(3.2.12)立即给出(3.2.11). ∎