第三章 无约束优化方法

第3.2节 最速下降法

天津大学数学学院

2020年3月2日

考虑无约束优化问题

$$\min_{x\in R^n} \ f(x) \qquad (3.0.1)$$

其中 $f:\mathbb{R}^n \to \mathbb{R}.$

记号： $g(x) := \nabla f(x), \quad g_k := \nabla f(x^k), \quad g_* := \nabla f(x^*)$

$G(x) := \nabla^2 f(x), \quad G_k := \nabla^2 f(x^k), \quad G_* := \nabla^2 f(x^*)$

最速下降方向

结论：设目标函数 f 在 x^k 附近连续可微，且 $g_k \triangleq \nabla f(x^k) \neq 0$，则 $-g_k$ 是在 f 在 x^k 处的最速下降方向.

证明：将 f 在 x^k 处Taylor展开，

$$f(x) = f(x^k) + g_k^T (x - x^k) + o(\| x - x^k \|).$$

记 $x - x^k = \lambda d^k$，则

$$f(x^k + \lambda d^k) - f(x^k) = \lambda g_k^T d^k + o(\| \lambda d^k \|).$$

若 d^k 满足 $g_k^T d^k < 0$，则 d^k 是下降方向，当 λ 取定后，$g_k^T d^k$ 的值越小，即 $-g_k^T d^k$ 的值越大，函数 f 在 x^k 处下降量越大.由Cauchy-Schwartz不等式，

$$\left| g_k^T d^k \right| \leq \| g_k \| \cdot \| d^k \|$$

可知当且仅当 $d^k = -g_k$ 时，$g_k^T d^k$ 最小，$-g_k^T d^k$ 最大，从而 $-g_k$ 是最速下降方向.

以$-g_k$作为下降方向的算法叫最速下降法，又称为梯度法.

最速下降法（steepest descent method）是一种以负梯度方向作为下降方向的极小化算法.

步长由精确一维线搜索确定的最速下降法是由Cauchy在1874年提出，它是求解无约束优化问题最早的数值方法.

最速下降法算法框架

算法3.2.1 (最速下降算法)

选择初始点$x^0 \in \mathbb{R}^n$, 令$k := 0$

步1: 计算$d^k = -g_k$;若g_k=0, 算法终止.

步2: 由线搜索确定步长$\lambda_k > 0$, 使得$f(x^k + \lambda_k d^k) < f(x^k)$.

步3: 置$x^{k+1} := x^k + \lambda_k d^k$, $k := k+1$, 转步1.

注 在理论分析中, 取g_k=0作为算法终止规则;

在实际计算中, 中止规则为$\|g_k\| < \varepsilon$, $(0 \le \varepsilon \le 1$是容许误差).

注 最速下降法的迭代格式为 $x^{k+1} = x^k - \lambda_k g_k$

回顾

线搜索迭代下降算法的收敛性

定理3.1.4 假设 f：$\mathbb{R}^n \to \mathbb{R}$连续可微有下界，且其梯度函数$g$具有Lipschitz连续性，即存在常数$L>0$使得对任意$x,y\in\mathbb{R}^n$,有$\|g(x)-g(y)\|\le L\|x-y\|$.

设无穷序列$\{x^k\}$由算法3.1.1产生, 其中步长λ_k由精确一维搜索得到，那么,

$$\sum_{k=0}^{\infty}\|g_k\|^2\ \cos^2\theta_k<\infty \quad (这里\theta_k表示d^k与-g_k之间的夹角).$$

特别地，若存在常数$\beta>0$使得$\cos\theta_k\ge\beta$,则$\lim\limits_{k\to\infty}\|g_k\|=0$.

定理3.1.5 假设无穷序列$\{x^k\}$由算法3.1.1产生，其中步长λ_k由Armijo线搜索得到，并且定理3.1.4的其他条件成立. 如果存在常数$c>0$, 使得

$$\|g_k\| \le c\|d^k\|,$$

则定理3.1.4的结论成立.

定理3.1.6 如果无穷序列$\{x^k\}$由算法3.1.1产生，其中步长λ_k由$Wolfe$线搜索得到，并且定理3.1.4的其他条件成立，则定理3.1.4的结论成立.

最速下降法的全局收敛性

定理3.2.1 (全局收敛性）

假设 $f:\mathbb{R}^n \to \mathbb{R}$连续可微且有下界，且梯度函数$g$具有Lipschitz连续性. 设序列$\{x^k\}$由最速下降算法3.2.1产生，其中步长$\lambda_k$由精确一维线搜索得到、或由Armijo线搜索得到、或由Wolfe线搜索得到. 则算法3.2.1或者有限终止，或者$\lim\limits_{k\to\infty}\|g_k\|=0$.

证明：定理3.1.4、3.1.5和3.1.6，取θ=0的情形.

最速下降法的优、缺点

优点： 程序设计简单，计算工作量小、存储量小；

对初始点没有特别要求，具有全局收敛性.

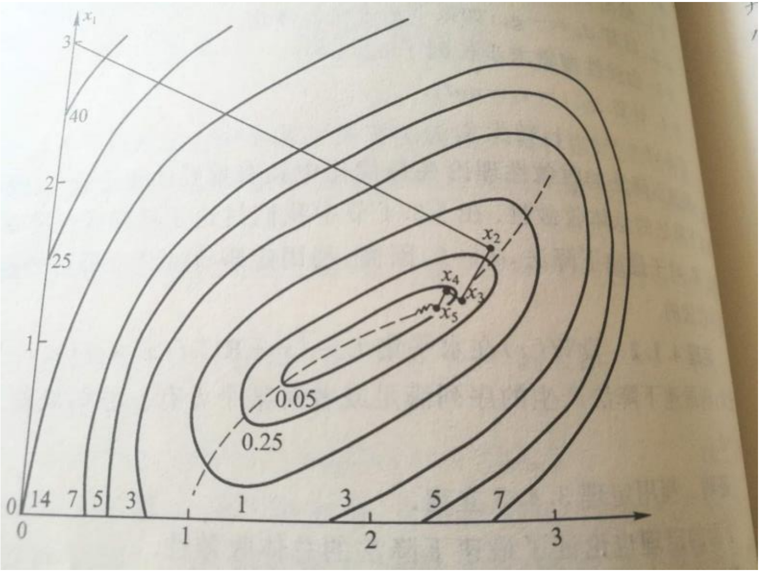
缺点：收敛速度慢，收敛速度是线性的（当步长由精确一维线搜索得到时）.

采用精确一维线搜索的最速下降法最多具有线性收敛速度，这与“最速下降方向”并不矛盾，最速下降方向只是反映了目标函数在迭代点的局部性质，收敛速度是从全局考虑的.

精确一维线搜索满足$g_{k+1}^{T}d^{k}=0$,
采用精确一维线搜索的最速下降法满足

$$g_{k+1}^{T}g_{k}=d_{k+1}^{T}d^{k}=0,$$

这表明最速下降法中相继两次的迭代方向是正交的，逼近极小点的路线是锯齿形，越接近极小点，步长越小，前进越慢.



例1 设正定二次函数 $f(x)=\frac{1}{2}x^TGx+q^Tx+r$，其中$G\in\mathbb{R}^{n\times n}$正定矩阵，$q\in\mathbb{R}^n$，$r\in\mathbb{R}$. 证明：由精确一维线搜索确定的步长$\lambda_k=-\frac{(g_k)^Td^k}{(d^k)^TG\,d^k}$.

证明 $g(x)=Gx+q, G(x)=G$，其中$G\in\mathbb{R}^{n\times n}$正定.

由精确一维线搜索确定的步长$\lambda_k=\arg\min\limits_{\lambda>0}\varphi(\lambda)=f(x^k+\lambda d^k)$,满足

$$0=\varphi'(\lambda_k)=\nabla f(x^k+\lambda_k d^k)^{\mathrm{T}}d^k=(G(x^k+\lambda_k d^k)+q)^Td^k$$

$$=(Gx^k+q)^Td^k+\lambda_k(d^k)^TGd^k=g_k^Td^k+\lambda_k(d^k)^TGd^k,$$

所以，$\lambda_k=-\frac{(g_k)^Td^k}{(d^k)^TG\,d^k}$.

注1 特别地, 若$d^k = -g_k$, 则 $\lambda_k = \dfrac{(g_k)^T g_k}{(g_k)^T G g_k}$.

注2 利用采用精确一维线搜索的最速下降法求解目标函数为正定二次函数的无约束极小化问题，迭代格式为：

$$x^{k+1} = x^k - \frac{(g_k)^T g_k}{(g_k)^T G g_k} g_k.$$

例3.2.1 利用采用精确一维线搜索的的最速下降算法求解

$$\min \quad f(x) = \frac{1}{2}x_1^2 + x_2^2,$$

其中初始点 $x^0 = (2,1)^{\mathrm{T}}$.

解 $g(x) = (x_1, 2x_2)^{\mathrm{T}}, G(x) = \begin{bmatrix} 1 & 0 \\ 0 & 2 \end{bmatrix} \succ 0.$

$$g_0 = g(x^0) = (2,2)^{\mathrm{T}} \neq \mathbf{0},$$

$$x^1 = x^0 + \lambda_0 d^0 = x^0 - \frac{(g_0)^T g_0}{(g_0\ )^T G\ g_0} g_0 = \begin{bmatrix} 2 \\ 1 \end{bmatrix} - \frac{8}{12}\begin{bmatrix} 2 \\ 2 \end{bmatrix} = \frac{1}{3}\begin{bmatrix} 2 \\ -1 \end{bmatrix}.$$

$$g_1 = g(x^1) = \frac{1}{3}(2,-2)^{\mathrm{T}} \neq \mathbf{0},$$

$$x^2 = x^1 + \lambda_1 d^1 = x^1 - \frac{(g_1)^T g_1}{(g_1)^T G g_1} g_1 = \frac{1}{3}\begin{bmatrix} 2 \\ -1 \end{bmatrix} - \frac{8/9}{12/9} \cdot \frac{1}{3}\begin{bmatrix} 2 \\ -2 \end{bmatrix} = (\frac{1}{3})^2 \begin{bmatrix} 2 \\ (-1)^2 \end{bmatrix}.$$

类似计算并归纳可得迭代点列为

$$x^k = (\frac{1}{3})^k \begin{bmatrix} 2 \\ (-1)^k \end{bmatrix}, \quad k = 0, 1, 2, \ldots$$

当$k \to \infty$时，有$x^k \to x^* = (0,0)^T$.

谢谢观看！