第三章 无约束优化方法

无约束优化问题

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

其中 $f:\mathbb{R}^n \to \mathbb{R}.$

天津大学数学学院

2020年3月2日

教学内容

一．算法理论基础（ § 3.1）

二．最速下降法（ § 3.2）

三．牛顿法（ §3.3）

四．共轭梯度法（ § 3.4）

五．拟牛顿法（ § 3.5）

§3.1 算法理论基础

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

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

一、无约束优化问题的最优性条件

引理3.1.1 给定 $x^* \in \mathbb{R}^n$和任意的$\boldsymbol{d} \in \mathbb{R}^n \setminus \{0\}$, 则 x^*是 $f:\mathbb{R}^n \to \mathbb{R}$的(严格)局部极小点$\Leftrightarrow \lambda_*=0$是一元函数$\varphi(\lambda)=f(x^*+\lambda d)$的(严格)局部极小点.

证明： 显然，由$\varphi(\lambda)=f(x^*+\lambda d)$可得$\varphi(0)=f(x^*)$.

假设 x^* 是多元实值函数 f 的局部极小值点, 那么对任意靠近 x^* 的 $x \in \mathbb{R}^n$ 有 $f(x^*) \leqslant f(x)$. 而对任意的 $d \in \mathbb{R}^n \setminus \{0\}$, 当 λ 充分靠近 0 时, $x^*+\lambda d$ 将充分靠近 x^*, 所以 $\lambda_*=0$ 是一元函数 $\varphi(\lambda)=f(x^*+\lambda d)$ 的局部极小值点.

假设 $\lambda_*=0$ 是一元函数 $\varphi(\lambda)=f(x^*+\lambda d)$ 的局部极小值点, 那么当 λ 充分靠近 0 时, $\varphi(0)\leqslant\varphi(\lambda)$. 因此, 对任意的 $d\in\mathbb{R}^n\setminus\{0\}$, 当 λ 充分靠近 0 时, $x^*+\lambda d$ 将充分靠近 x^* 且 $f(x^*)\leqslant f(x^*+\lambda d)$. 由 d 的任意性可得, 对任意靠近 x^* 的 $x\in\mathbb{R}^n$ 有 $f(x^*)\leqslant f(x)$. 所以 x^* 是多元实值函数 f 的局部极小值点.

因此, x^* 是多元实值函数 f 的局部极小值点当且仅当 $\lambda_*=0$ 是一元函数 $\varphi(\lambda)=f(x^*+\lambda d)$ 的局部极小值点. 类似可得: x^* 是多元实值函数 f 的严格局部极小值点当且仅当 $\lambda_*=0$ 是一元函数 $\varphi(\lambda)=f(x^*+\lambda d)$ 的严格局部极小值点. □

一元函数取得极值的条件

设一元函数$\varphi:\mathbb{R}\to\mathbb{R}$，则

(i)（一阶必要条件）若函数φ一阶连续可微且$\lambda_*\in\mathbb{R}$为φ的局部极小值点，则$\varphi'(\lambda_*)=0$;

(ii)（二阶必要条件）若函数φ二阶连续可微且$\lambda_*\in\mathbb{R}$为φ的局部极小值点，则$\varphi'(\lambda_*)=0$且$\varphi''(\lambda_*)\geq 0$;

(iii)（二阶充分条件）若函数φ二阶连续可微且$\varphi'(\lambda_*)=0$且$\varphi''(\lambda_*)>0$,则$\lambda_*\in\mathbb{R}$为函数φ的严格局部极小值点.

无约束优化问题的最优性条件

定理3.1.1 (必要条件) 设 f：$\mathbb{R}^n \to \mathbb{R}$.

(i) （一阶必要条件）若f在$x^* \in \mathbb{R}^n$的某邻域内一阶连续可微，且x^*是f的局部极小点，则 $g_* = \nabla f(x^*) = 0$；

(ii) （二阶必要条件）若f在$x^* \in \mathbb{R}^n$的某邻域内二阶连续可微，且x^*是f的局部极小点，则 $g_* = 0$ 且 $G_* = \nabla^2 f(x^*)$ 对称半正定.

称满足一阶必要条件 $g_* = 0$ 的点为稳定点（也称为驻点）.

稳定点分为三种类型：极大值点、极小值点、鞍点.

注：该定理的逆命题不成立，可能有鞍点存在.

证明 只证明 (ii) 中的结论, (i) 中的结论类似可得. 任取 $d \in \mathbb{R}^n \setminus \{0\}$, 定义 $\varphi(\lambda) = f(x^* + \lambda d)$. 由于函数 f 在 x^* 的某邻域内二阶连续可微, 所以函数 φ 在 $\lambda_* = 0$ 的某邻域内二阶连续可微且

$$\varphi'(\lambda) = g(x^* + \lambda d)^{\mathrm{T}} d, \quad \varphi''(\lambda) = d^{\mathrm{T}} G(x^* + \lambda d) d.$$

由于 x^* 为 f 的局部极小值点, 所以由引理 3.1.1 知, $\lambda_* = 0$ 是一元函数 φ 的局部极小值点. 因此, 由一元函数的极值条件 $\varphi'(0) = 0$ 和 $\varphi''(0) \geqslant 0$ 可得

$$g_*^{\mathrm{T}} d = 0, \quad d^{\mathrm{T}} G_* d \geqslant 0.$$

由 d 的任意性进一步可得: $g_* = 0$ 且 G_* 半正定. 定理得证. □

定理3.1.2 (二阶充分条件)

设 f 在 $x^* \in \mathbb{R}^n$ 的某邻域内二阶连续可微. 若 $g_* = 0$ 且 G_* 正定，则 x^* 是函数 f 的严格局部极小值点.

证明 任取 $d \in \mathbb{R}^n \setminus \{0\}$, 定义 $\varphi(\lambda) = f(x^* + \lambda d)$. 由于函数 f 在 x^* 的某邻域内二阶连续可微, 所以函数 φ 在 $\lambda_* = 0$ 的某邻域内二阶连续可微且

$$\varphi'(0) = g_*^{\mathrm{T}} d, \quad \varphi''(0) = d^{\mathrm{T}} G_* d.$$

由于 $g_* = 0$ 且 G_* 正定, 所以可得

$$\varphi'(0) = g_*^{\mathrm{T}} d = 0 \quad 且 \quad \varphi''(0) = d^{\mathrm{T}} G_* d > 0, \quad \forall d \in \mathbb{R}^n \setminus \{0\}.$$

这表明: $\lambda_* = 0$ 是一元函数 φ 的严格局部极小值点. 所以由引理 3.1.1 可知, x^* 为 f 的严格局部极小值点. 定理得证. $\square$

无约束凸规划问题的最优性条件

定理3.1.3 (无约束凸规划的最优性条件)

设 $f:\mathbb{R}^n \to \mathbb{R}$是一阶连续可微的凸函数，则 x^*是函数 f的全局极小值点 $\Leftrightarrow g_* = 0$.

证明 必要性由定理 3.1.1 可得, 下证充分性. 由于 f 是 $\mathbb{R}^n$ 上的一阶连续可微的凸函数, 根据凸函数的一阶判别定理 (定理 1.3.4), 因而可得对任意的 $x \in \mathbb{R}^n$, 有

$$f(x) - f(x^*) \geqslant g_*^{\mathrm{T}}(x - x^*).$$

结合上式和条件 $g_* = 0$, 于是有: 对任意的 $x \in \mathbb{R}^n$ 有 $f(x) \geqslant f(x^*)$. 这表明: x^* 为函数 f 的全局极小值点. 定理得证. □

二、线搜索迭代下降算法及其收敛性

1、算法的基本思想

从某个初始点开始，按照一定的规则不断地寻找迭代方向和迭代步长进行迭代，使得目标函数 f 的函数值不断地下降，以至于所得到的点列中的最后一个点（如果点列有限）或点列的某个聚点（如果点列无限）是无约束优化问题的稳定点.

二、线搜索迭代下降算法及其收敛性

2、算法框架

算法3.1.1 (线搜索迭代下降算法）选择初始点$x^0 \in \mathbb{R}^n$,置$k := 0$.

步1: 若$g_k = g(x^k) = 0$,则算法终止.

步2: 确定下降方向$d^k \in \mathbb{R}^n$,使得$g_k^T d^k < 0$.

步3: 确定步长$\lambda_k > 0$,使得$f(x^k + \lambda_k d^k) < f(x^k)$.

步4: 置$x^{k+1} := x^k + \lambda_k d^k$, $k := k+1$, 并返回步1.

3、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,

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

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

证明 由于最后的结论可由(3.1.2)式的结论直接得到，因而下面只需证明(3.1.2)式的结论成立. 因为函数f连续可微, 所以对任意的$\lambda > 0$, 存在$\alpha_k \in (0,1)$使得

$$
\begin{aligned}
f(x^k + \lambda d^k) &= f(x^k) + \lambda g(x^k + \alpha_k \lambda d^k)^\top d^k \\
&= f(x^k) + \lambda g_k^\top d^k + \lambda [g(x^k + \alpha_k \lambda d^k) - g_k]^\top d^k \\
&\le f(x^k) + \lambda g_k^\top d^k + \lambda \|g(x^k + \alpha_k \lambda d^k) - g_k\| \|d^k\| \\
&< f(x^k) + \lambda g_k^\top d^k + \lambda^2 \mathcal{L} \|d^k\|^2,
\end{aligned}
$$

特别地, 取$\hat{\lambda}_k := -\frac{g_k^\top d^k}{2\mathcal{L}\|d^k\|^2}$, 则由上式可得

$$
f(x^k + \hat{\lambda}_k d^k) \le f(x^k) - \frac{(g_k^\top d^k)^2}{4\mathcal{L}\|d^k\|^2}.
$$

又由精确一维线搜索知：步长λ_k满足$f(x^{k+1}) = f(x^k + \lambda_k d^k) \le f(x^k + \hat{\lambda}_k d^k)$. 因此，

$$f(x^{k+1}) \le f(x^k + \hat{\lambda}_k d^k) \le f(x^k) - \frac{(g_k^\top d^k)^2}{4\mathcal{L}\|d^k\|^2}.$$

根据(3.1.1)式, 由上式可得

$$\|g_k\|^2 \cos^2\theta_k \le 4\mathcal{L}[f(x^k) - f(x^{k+1})].$$

从而有 $\sum_{k=0}^{n-1}\|g_k\|^2 \cos^2\theta_k \le 4L\sum_{k=0}^{n-1}(f(x^k) - f(x^{k+1})) = 4L(f(x^0) - f(x^n))$

因为$f(x)$有下界，所以$\sum_{k=0}^{\infty}\|g_k\|^2 \cos^2\theta_k < \infty$. 证毕.

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.

4、算法收敛速度

定理3.1.7 假设f：$\mathbb{R}^n \to \mathbb{R}$二阶连续可微, 无穷序列$\{x^k\}$由算法3.1.1产生，其中迭代步长$\lambda_k$由Armijo线搜索得到(其中$\sigma \in (0,\frac{1}{2})$)或由Wolfe线搜索得到. 假设$x^k \to x^*$， $g_* = 0$且G_*正定. 如果

$$\lim_{k\to\infty} \frac{\left\| g_k + G_k d^k \right\|}{\left\| d^k \right\|} = 0,$$

则 (1) 当k充分大时，$\lambda_k = 1$; (2) 序列$\{x^k\}$局部超线性收敛于x^*.

谢谢观看！