最优化理论与方法

第10讲

下降算法的收敛速度

下降算法的一般步骤

算法 2.1 *(求解无约束问题的下降算法)*

步 *1* 给定初始点 $x^{(0)} \in R^n$, 精度 $\epsilon > 0$. 令 $k := 0$.

步 *2* 若 $\|\nabla f(x^{(k)})\| \leq \epsilon$, 则终止算法，得解 $x^{(k)}$. 否则，转步 *3*.

步 *3* 确定下降方向 $d^{(k)}$, 使得

$$\nabla f(x^{(k)})^T d^{(k)} < 0.$$

步 *4* 确定步长 $\alpha_k > 0$ 使得

$$f(x^{(k)} + \alpha_k d^{(k)}) < f(x^{(k)}).$$

步 5 令 $x^{(k+1)} = x^{(k)} + \alpha_k d^{(k)}$, $k := k + 1$. 转步 *2*.

$$\boxed{x^{k+1} = x^k + \alpha^k d^k, \qquad k = 0, 1, \ldots}$$

步长 α^k 的选取：线性搜索

计算步长α_k的两种线性搜索：精确搜索和非精确搜索：

精确搜索：α_k是一维优化问题的解

$$\min_{\alpha>0} f(x_k + \alpha d_k)$$

即α_k满足

$$\nabla f(x_k + \alpha_k d_k)^{\mathrm{T}} d_k = 0$$

非精确搜索：α_k按照某种规则计算使之满足

$$f(x_k + \alpha_k d_k) < f(x_k)$$

即函数值有一定程度下降.

二、 非精确线性搜索— Armijo 型线性搜索和 Wolfe-Powell 型线性搜索

Armijo型线性搜索：给定 $\sigma_1 \in (0,1)$, 计算 α_k 满足

$$f(x_k + \alpha_k d_k) \le f(x_k) + \sigma_1 \alpha_k \nabla f(x_k)^{\mathrm{T}} d_k \qquad (2.6)$$

令 $\phi(\alpha) = f(x_k + \alpha d_k)$, (2.6)等价于

$$\phi(\alpha_{\mathrm{k}}) \le \phi(0) + \sigma_1 \alpha_k \phi^{'}(0)$$

Wolfe - Powell 线性搜索：给定常数 $0 < \sigma_1 < \sigma_2 < 1$

$$\begin{cases} f(x_k + \alpha_k d_k) \le f(x_k) + \sigma_1 \alpha_k \nabla f(x_k)^T d_k \\ \nabla f(x_k + \alpha_k d_k)^{\mathrm{T}} d_k \ge \sigma_2 \nabla f(x_k)^T d_k \end{cases} \qquad (2.7)$$

四、下降算法的全局收敛性和超线性收敛性

现在我们来研究下降算法2.1的全局收敛性和超线性收敛速度.

记向量d_k和$-\nabla f(x_k)$的夹角为θ_k,则有

$$\cos\theta_k = \frac{-\nabla f(x_k)^{\mathrm{T}} d_k}{\|\nabla f(x_k)\|\,\|d_k\|} \qquad (2.12)$$

从下面的定理可以看到，下降算法的收敛性与夹角θ_k有重要的关系:

首先我们给出下面的基本假设:

假设2.4.1 函数$f(x)$连续可微有下界且$\nabla f(x)$Lipschitz 连续,即存在常数$L>0$,使得

$$\|\nabla f(x) - \nabla f(y)\| \le L\|x-y\|, \quad \forall x, y \in R^n$$

定理2.4.1-2.4.2 设假设2.4.1成立，序列$\{x_k\}$由算法2.1产生，其中步长α_k由精确搜索或Wolfe-Powell搜索产生，则

$$\sum_{k=0}^{\infty} \|\nabla f(x_k)\|^2 \cos^2\theta_k < +\infty \qquad (2.13)$$

特别地，若存在常数$\delta > 0$使得$\cos\theta_k \geq \delta$，则

$$\lim_{k\to\infty} \|\nabla f(x_k)\| = 0 \qquad (2.14)$$

如果(2.14)成立，我们认为算法是全局收敛的.

该定理表明：算法的收敛性与夹角有关

由于Wolfe-Powell包含精确搜索，我们只需证明定理对Wolfe-Powell搜索成立

定理2.4.3 设假设2.4.1成立, 序列$\{x_k\}$由算法2.1产生, 其中步长α_k由Armijo 搜索产生, 且存在常数$C>0$, 使得

$$\|\nabla f(x_k)\| \le C\|d_k\| \tag{2.17}$$

则定理2.4.1的结论成立

定理2.4.4 设假设2.4.1成立,序列$\{x_k\}$由算法2.1产生,其中步长α_k由精确搜索或Wolfe-Powell搜索或Armijo搜索确定且(2.17)成立. 若进一步

$$\liminf_{k\to\infty} \cos\theta_k > 0 \tag{2.18}$$

则

$$\liminf_{k\to\infty} \|\nabla f(x_k)\| = 0 \tag{2.19}$$

特别地,（2.19）成立，若存在常数$\eta>0$,使得

$$\prod_{i=0}^{k-1} \cos\theta_i \ge \eta^k \tag{2.20}$$

全局收敛性定理含义

1) 算法产生的点列收敛于问题的稳定点（或解）：

$$\lim_{k\to\infty} x^{(k)} = x^*, \quad \nabla f(x^*) = 0.$$

2) 算法产生的点列的每一个极限点都是问题的稳定点（或解）：

$$\lim_{k\to\infty} \|\nabla f(x^{(k)})\| = 0.$$

3) 算法产生的点列中有一个极限点是问题的稳定点（或解）：

$$\liminf_{k\to\infty} \|\nabla f(x^{(k)})\| = 0.$$

$$1) \quad \Rightarrow \quad 2) \quad \Rightarrow \quad 3)$$

三种全局收敛性的等价条件

定理（三种全局收敛性的等价条件）

设$\{x^{(k)}\}$是下降算法产生的点列. 若f是一致凸函数，则三种收敛性等价. 即由

$$\liminf_{k\to\infty} \|\nabla f(x^{(k)})\| = 0.$$

可推得：$\{x^{(k)}\}$收敛于函数f的唯一全局最小值点.

证明： 由条件知：存在子列$\{x^{(k)}\}_K \to x^*, \nabla f(x^*) = 0$. 由于$f$是一致凸函数，因此$x^*$是问题的全局最小值点.

另一方面, 函数值序列$\{f(x^{(k)})\}$单调递减，因此$\{f(x^{(k)})\} \to f(x^*)$. 从而，$\{x^{(k)}\}$的每一个极限点处的函数值都是$f(x^*)$，即$f$的全局最小值. 即$\{x^{(k)}\}$的每一个极限点都是$f$的全局最小值点. 但$f$的全局最小值点唯一，因此，$\{x^{(k)}\} \to x^*$. □

收敛速度

设算法产生的点列$\{x^{(k)}\}\subset R^n$收敛于点x^*.

(1) q-线性收敛：存在常数$\rho\in(0,1)$使得当k充分大时，

$$\|x^{(k+1)}-x^*\|\leq\rho\|x^{(k)}-x^*\|,$$

r-线性收敛：若$\|x^{(k+1)}-x^*\|\leq r_k$，且$\{r_k\}$具有线性收敛速度.

(2) 超线性收敛：若

$$\lim_{k\to\infty}\frac{\|x^{(k+1)}-x^*\|}{\|x^{(k)}-x^*\|}=0,$$

则称$\{x^{(k)}\}$超线性收敛于x^*. 或称$\{x^{(k)}\}$的收敛速度是超线性的.

(3) 二次收敛：若存在常数$M>0$使得当k充分大时，

$$\|x^{(k+1)}-x^*\|\leq M\|x^{(k)}-x^*\|^2,$$

则称$\{x^{(k)}\}$二次收敛于x^*. 或称$\{x^{(k)}\}$的收敛速度是二次的.

关于下降算法的收敛速度，我们有

定理2.5.1　设函数f二次连续可微，点列$\{x_k\}$由算法2.1 产生，$\{x_k\} \to x^*$,且$\nabla f(x^*) = 0, \nabla^2 f(x^*)$正定, 其中步长$\alpha_k$由精确搜索或Wolfe-Powell搜索, 或Armijo线性搜索（且（2.17）成立）确定，若存在常数$\eta > 0$, 使得$\prod_{i=0}^{k-1} \cos\theta_i \geq \eta^k$,则存在常数b>0, r$\in(0,1)$使得当k充分大时，

$$\|x^{(k+1)} - x^*\| \leq br^k \qquad (2.24)$$

进一步关于下降算法的收敛速度，我们有

定理2.5.2 设函数f二次连续可微，点列$\{x_k\}$由算法2.1产生其中步长α_k由Armijo搜索或Wolfe - Powell搜索确定，其中$\sigma_1 \in (0,1/2)$. 设$\{x_k\} \to x^*$，且$\nabla f(x^*) = 0, \nabla^2 f(x^*)$正定. 若

$$\lim_{k\to\infty} \frac{\| \nabla f(x_k) + \nabla^2 f(x_k) d_k \|}{\| d_k \|} = 0$$

则

(1) 当k充分大时，$\alpha_k = 1$;

(2) 序列$\{x_k\}$超线性收敛于x^*;

(3) 若$\nabla^2 f(x)$在x^*处Lipschitz连续，且

$$\delta_k = \frac{\| \nabla f(x_k) + \nabla^2 f(x_k) d_k \|}{\| d_k \|} = O(\| x_k - x^* \|)$$

则x_k二次收敛于x^*.

上述定理说明了下降算法的收敛速度 和取单位步长的重要性

超线性收敛性定理：证明步骤

Dennis-Moré 条件：

$$\lim_{k\to\infty} \frac{\|\nabla f(x^{(k)}) + \nabla^2 f(x^{(k)}) d^{(k)}\|}{\|d^{(k)}\|} = 0.$$

- 第一步：存在常数$m_1 > 0$使得不等式

$$\|d^{(k)}\| \le m_1 \|\nabla f(x^{(k)})\| = O(\|x^{(k)} - x^*\|)$$

  对所有充分大的k均成立. 特别，$\{d^{(k)}\} \to 0.$
- 第二步：存在常数$\eta > 0$，使得当k充分大时，

$$-\nabla f(x^{(k)})^T d^{(k)} \ge \eta \|d^{(k)}\|^2.$$

- 第三步：若$\alpha_k = 1$, 则算法具有超线性收敛性速度. 即

$$\|x^{(k)} + d^{(k)} - x^*\| = o(\|x^{(k)} - x^*\|).$$

- 第四步：$\alpha_k = 1$满足Armijo和Wolfe-Powell 线性搜索的条件.

超线性收敛性定理：证明第一步

第一步：存在常数$m_1 > 0$使得不等式

$$\|d^{(k)}\| \le m_1 \|\nabla f(x^{(k)})\| = O(\|x^{(k)} - x^*\|) \tag{13}$$

对所有充分大的k均成立. 特别，$\{d^{(k)}\} \to 0$.
由于$\{x^{(k)}\} \to x^*$且$\nabla^2 f(x^*)$正定，故当k充分大时，$\nabla^2 f(x^{(k)})$一致正定. 故存在常数$m > 0$ 使得不等式

$$\|\nabla^2 f(x^{(k)}) d\| \ge m\|d\|$$

对所有$d \in R^n$以及所有充分大的k都成立. 令

$$\delta_k = \frac{\|\nabla f(x^{(k)}) + \nabla^2 f(x^{(k)}) d^{(k)}\|}{\|d^{(k)}\|}.$$

则当k充分大时，$\delta_k < m/2$. 而且，

$$m\|d^{(k)}\| \le \|\nabla^2 f(x^{(k)}) d^{(k)}\| \le \delta_k \|d^{(k)}\| + \|\nabla f(x^{(k)})\|.$$

因此，存在常数$m_1 > 0$使得(13)对所有充分大的k均成立.

超线性收敛性定理：证明第二步

第二步：存在常数$\eta > 0$，使得当k充分大时，

$$-\nabla f(x^{(k)})^T d^{(k)} \geq \eta \|d^{(k)}\|^2.$$

我们有

$$\begin{aligned}
-\nabla f(x^{(k)})^T d^{(k)} &= d^{(k)T} \nabla^2 f(x^k) d^{(k)} - d^{(k)T} [\nabla f(x^{(k)}) + \nabla^2 f(x^{(k)}) d^{(k)}] \\
&\geq m\|d^{(k)}\|^2 - \|d^{(k)}\| \, \|\nabla f(x^{(k)}) + \nabla^2 f(x^{(k)}) d^{(k)}\| \\
&= (m - \delta_k)\|d^{(k)}\|^2.
\end{aligned}$$

由此及(12)可得，存在常数$\eta > 0$，使得当k充分大时，

$$-\nabla f(x^{(k)})^T d^{(k)} \geq \eta \|d^{(k)}\|^2.$$

超线性收敛性定理：证明第三步

第三步：若$\alpha_k = 1$, 则算法具有超线性收敛性速度.
由$\nabla^2 f(x_k)$的一致正定性得

$$
\begin{aligned}
m\|x^{(k+1)} - x^*\| &\leq \|\nabla^2 f(x^{(k)})(x^{(k)} + d^{(k)} - x^*)\| \\
&= \|\nabla^2 f(x^{(k)})(x^{(k)} - x^*) - \nabla f(x^{(k)}) + \nabla^2 f(x^{(k)})d^{(k)} + \nabla f(x^{(k)})\| \\
&\leq \|\nabla f(x^{(k)}) - \nabla f(x^*) - \nabla^2 f(x^{(k)})(x^{(k)} - x^*)\| \\
&\quad + \|\nabla^2 f(x^{(k)})d^{(k)} + \nabla f(x^{(k)})\| \\
&\leq o(\|x^{(k)} - x^*\| + o(\|d^{(k)}\|) = o(\|x^{(k)} - x^*\|),
\end{aligned}
$$

其中，最后一个不等式由(12)得到，最后一个等式由(13)得到. 上面的不等式说明$\{x^{(k)}\}$超线性收敛.

超线性收敛性定理：证明第四步

第四步：$\alpha_k = 1$满足Armijo和Wolfe-Powell 线性搜索的条件.

(1) 先证明$\alpha_k = 1$满足Armijo线性搜索条件. 利用Taylor展开，有

$$\begin{aligned}
& f(x^{(k)} + d^{(k)}) - f(x^{(k)}) - \sigma_1 \nabla f(x^{(k)})^T d^{(k)} \\
= \;& [f(x^{(k)} + d^{(k)}) - f(x^{(k)}) - \frac{1}{2}\nabla f(x^{(k)})^T d^{(k)}] + (\frac{1}{2} - \sigma_1)\nabla f(x^{(k)})^T d^{(k)} \\
= \;& (\frac{1}{2} - \sigma_1)\nabla f(x^{(k)})^T d^{(k)} + \frac{1}{2} d^{(k)T}[\nabla^2 f(x^{(k)}) d^{(k)} + \nabla f(x^{(k)})] + o(\|d^{(k)}\|^2) \\
= \;& (\frac{1}{2} - \sigma_1)\nabla f(x^{(k)})^T d^{(k)} + o(\|d^{(k)}\|^2) \\
\leq \;& (\sigma_1 - \frac{1}{2})\eta \|d^{(k)}\|^2 + o(\|d^{(k)}\|^2).
\end{aligned}$$

因此，当k充分大时, $f(x^{(k)} + d^{(k)}) - f(x^{(k)}) - \sigma_1 \nabla f(x^{(k)})^T d^{(k)} \leq 0$，即当$k$充分大时，$\alpha_k = 1$满足Armijo型线性搜索的条件以及Wolfe-Powell型线性搜索中的第一个不等式.

超线性收敛性定理：证明第四步

再证明$\alpha_k = 1$满足Wolfe-Powell型线性搜索中的第二个不等式.
利用中值定理得

$$
\begin{aligned}
&\nabla f(x^{(k)} + d^{(k)})^T d^{(k)} - \sigma_2 \nabla f(x^{(k)})^T d^{(k)} \\
&= [\nabla f(x^{(k)} + d^{(k)}) - \nabla f(x^{(k)})]^T d^{(k)} + (1 - \sigma_2)\nabla f(x^{(k)})^T d^{(k)} \\
&= d^{(k)T} \nabla^2 f(x^{(k)}) d^{(k)} + (1 - \sigma_2)\nabla f(x^{(k)})^T d^{(k)} + o(\|d^{(k)}\|^2) \\
&= -\sigma_2 \nabla f(x^{(k)})^T d^{(k)} + d^{(k)T}[\nabla f(x^{(k)}) + \nabla^2 f(x^{(k)})] d^{(k)} + o(\|d^{(k)}\|^2) \\
&\geq \eta \sigma_2 \|d^{(k)}\|^2 + o(\|d^{(k)}\|^2).
\end{aligned}
$$

因此，当k充分大时, $\nabla f(x^{(k)} + d^{(k)})^T d^{(k)} - \sigma_2 \nabla f(x^{(k)})^T d^{(k)} \geq 0$，即当$k$充分大时, $\alpha_k = 1$满足Wolfe-Powell型线性搜索中的第二个不等式.

作 业

- 习题2:

  17, 20

课后：复习掌握下降算法的收敛性和收敛速度定理及其证明。