3.7 信赖域方法

前面几节介绍的求解无约束优化问题的方法，都为线搜索方法，其基本思想为：在每一步迭代中，得到迭代点x^k和迭代方向d^k后，从x^k出发沿d^k需要进行一维线搜索得到迭代步长λ_k，然后按$x^{k+1}=x^k+\lambda_k d^k$ 进行迭代，得到下一步的迭代点x^{k+1}. 与线搜索方法不同，信赖域方法的基本思想为: 在得到迭代点x^k后，确定一个合适的变化范围(通常取为以x^k为中心的球域)，称之为信赖域，在当前迭代点附近定义目标函数的一个近似二次模型，取二次模型在该信赖域内的极小值点作为下一次的迭代点，重复迭代直到算法终止. 在信赖域方法中，所考虑的二次模型在信赖域内的极小化问题被称为信赖域子问题.

3.7 信赖域方法

假设x^k为第k步迭代的迭代点，定义信赖域为$\{x \in \mathbb{R}^n \mid \|x - x^k\| \le r_k\}$, 其中$r_k$称为信赖域半径. 考虑函数$f$在迭代点$x^k$处的二阶Taylor展开, 并且令

$$f(x) \approx f(x^k) + g_k^\top (x - x^k) + \frac{1}{2}(x - x^k)^\top G_k (x - x^k).$$

若记$d := x - x^k$, 则信赖域子问题为

$$\begin{aligned} \min \quad & q_k(d) := f(x^k) + \tfrac{1}{2} g_k^\top d + d^\top B_k d \\ \text{s.t.} \quad & \|d\| \le r_k, \end{aligned} \tag{3.7.1}$$

其中B_k是一个对称矩阵，它是Hesse阵G_k或是G_k的一个近似矩阵. 显然，$q_k(0) = f(x^k)$.

3.7 信赖域方法

在信赖域方法中，如何选择信赖域半径是至关重要的.

- 如果信赖域半径太小，虽然二次模型与目标函数会更接近，但可能会失去下一个迭代点与目标函数的极小值点更靠近的机会，从而影响算法的效率；
- 如果信赖域半径太大，那么二次模型与目标函数的近似度差，从而使得二次模型的极小值点与目标函数的极小值点相距太远，以至于目标函数在新迭代点处的函数值得不到应有的下降.

3.7 信赖域方法

因此，在信赖域方法中，通常根据二次模型与目标函数的近似程度，按照一定的规则来选择合适的信赖域半径. 具体地，假设d^k是信赖域子问题(3.7.1)的最优解，利用目标函数定义实际下降量为$\mathrm{Ared}^k = f(x^k) - f(x^k + d^k)$; 且利用二次模型函数定义预测下降量为$\mathrm{Pred}^k = q_k(0) - q_k(d^k)$. 进一步，定义比值

$$\rho_k = \frac{\mathrm{Ared}^k}{\mathrm{Pred}^k} = \frac{f(x^k) - f(x^k + d^k)}{q_k(0) - q_k(d^k)}. \tag{3.7.2}$$

显然，ρ_k越接近1，函数f与函数q_k的一致性程度越好; $\rho_k \in [0, 1]$越接近0(甚至于ρ_k取负值)，函数f与函数q_k的一致性程度越差. 因此，若ρ_k接近1，可增大信赖域半径; 若$\rho_k \in [0, 1]$适中，保持信赖域半径不变; 若$\rho_k \in [0, 1]$接近0(或ρ_k取负值)，可减小信赖域半径. 具体的算法如下:

3.7 信赖域方法

算法 3.7.1 *(信赖域算法)* 选取初始点$x^0 \in \mathbb{R}^n$, 信赖域半径的上界$\hat{r}$, 初始信赖域半径$r_0 \in (0, \hat{r})$, 参数$0 < \eta_1 \leq \eta_2 < 1$, $0 < \gamma_1 < 1 < \gamma_2$. 置$k := 0$.

步*1* 若$g_k = 0$, 算法终止, 得到解x^k. 否则, 转步*2*.

步*2* 求解信赖域子问题*(3.7.1)*, 得到的解为d^k.

步*3* 由*(3.7.2)*式计算ρ_k. 如果$\rho_k < \eta_1$, 那么置$x^{k+1} := x^k$; 否则, 置$x^{k+1} := x^k + d^k$.

步*4* (信赖域半径的校正) 选择

$$\begin{cases} r_{k+1} \in (0, \gamma_1 r_k], & \text{若} \rho_k < \eta_1, \\ r_{k+1} \in [\gamma_1 r_k, r_k], & \text{若} \rho_k \in [\eta_1, \eta_2), \\ r_{k+1} \in [r_k, \min\{\gamma_2 r_k, \hat{r}\}], & \text{若} \rho_k \geq \eta_2 \text{且} \|d^k\| = r_k. \end{cases}$$

步*5* 产生B_{k+1}, 校正q_k. 置$k := k + 1$, 转步*1*.

3.7 信赖域方法

注. (i) 在算法3.7.1中, 参数通常取$\eta_1 = 1/4, \eta_2 = 3/4, \gamma_1 = 1/2$, 以及$\gamma_2 = 2$; 初始信赖域半径通常取为$r_0 = 1$或$r_0 = \|g_0\|/10$. (ii) 信赖域半径的具体取法之一如下: 若$\rho_k < \eta_1$, 则取$r_{k+1} = \gamma_1 r_k$; 若$\rho_k \in [\eta_1, \eta_2)$, 则取$r_{k+1} = r_k$; 若$\rho_k \geq \eta_2$且$\|d^k\| = r_k$, 则取$r_{k+1} = \min\{\gamma_2 r_k, \hat{r}\}$. (iii) 若$\rho_k \geq \eta_1$, 则称该步为成功步; 否则称其为不成功步.

在一定的条件下, 算法3.7.1具有全局收敛性. 下面的定理是在子问题(3.7.1)精确求解的情况下讨论算法3.7.1的收敛性.

定理 3.7.1 *假设算法3.7.1中的子问题(3.7.1)精确求解且当*$\rho_k \in [\eta_1, \eta_2)$*时取*$r_{k+1} = r_k$. *令序列*$\{x^k\}$*是由算法3.7.1产生的迭代序列. 假设水平集*$\mathscr{L}(f) = \{x \in \mathbb{R}^n \mid f(x) \leq f(x^0)\}$*是紧的, 且*$f(\cdot), \nabla f(\cdot)$*和*$G(\cdot)$*在*$\mathscr{L}(f)$*上连续. 那么算法3.7.1或者有限终止; 或者有*

$$\liminf_{k\to\infty} \|g_k\| = 0. \tag{3.7.3}$$