1.4 线搜索迭代算法概述及收敛性准则

1.4.1 线搜索算法的一般框架

基本思想：从某个点$x^0 \in R^n$出发，按照某种规则产生一个迭代点序列$\{x^k\}$,直到算法终止.

算法分类：

按照迭代点是否可行，分为

（1）可行算法 所有迭代点都是可行点

（2）不可行算法 迭代点中至少存在一个不可行点

按照目标函数值是否下降分为：

（1）单调下降算法 $f(x^{k+1}) \le f(x^k), \forall k$

(2) 非单调下降算法 否则

(一) 线搜索迭代算法的一般框架

步1：选择一个初始点$x^0 \in R^n$, 令$k := 0$;

步2： 判断当前的迭代点x^k是否满足终止条件；

步3： 从当前点出发，选择沿什么方向进行迭代，

（即迭代方向，不妨记为d^k)

以及沿该方向走多远（即迭代步长，不妨记为λ_k），

确定下一个迭代点$x^{k+1} = x^k + \lambda_k d^k$

步4： 令$k := k+1$, 转步2

注； 最有用的终止规则： $\left|f(x^k)-f(x^*)\right| \le \varepsilon; \left\|x^k - x^*\right\| \le \varepsilon, (x^* \in S)$

关于步2：常用的终止准则：

$(a)\quad \left\|x^{k+1}-x^k\right\| < \varepsilon$

$(b)\quad \left|f(x^{k+1})-f(x^k)\right| < \varepsilon$

$(c)\quad \dfrac{\left\|x^{k+1}-x^k\right\|}{\left\|x^k\right\|} < \varepsilon$

$(d)\quad \dfrac{\left|f(x^{k+1})-f(x^k)\right|}{\left|f(x^k)\right|} < \varepsilon$

$(e)\quad \left\|\nabla f(x^k)\right\| \le \varepsilon$

1.4.2 迭代方向

定义1.4.1 在点x^k处，对于向量$d^k \in R^n \setminus \{0\}$,若存在$\bar{\lambda} > 0$使得对任意$\lambda \in (0, \bar{\lambda})$都有$f(x^k + \lambda d^k) < f(x^k)$,则称$d^k$为函数$f$在$x^k$处的一个下降方向。

命题1.4.1 假设函数f一阶连续可微，那么d^k为f在x^k处的下降方向当且仅当$\nabla f(x^k)^T d^k < 0$.

证明：由Taylor展式得：

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

所以$f(x^k + \lambda d^k) - f(x^k) = \lambda[\nabla f(x^k)^T d^k + o(\|d^k\|)]$.

从而结论成立。

定义1.4.2 给定非空集合$F \subseteq R^n$和点$x^k \in F$.对于向量$d^k \in R^n \setminus \{0\}$.
若存在$\bar{\lambda} > 0$使得对任意$\lambda \in (0, \bar{\lambda})$都有$x^k + \lambda d^k \in F$,则称$d^k$为$x^k$处关于$F$的一个可行方向。

可行下降方向： 如果d^k为点x^k处的可行下降方向，那么存在λ_k使得
$x^k + \lambda_k d^k \in F$,且$f(x^k + \lambda_k d^k) < f(x^k)$.

1.4.3 迭代步长

1. 精确一维搜索

选取$\lambda_k = \arg\min_{\lambda>0} \varphi(\lambda) = \min_{\lambda>0} f(x^k + \lambda d^k)$,

$$f(x^{k+1}) = f(x^k + \lambda_k d^k)$$

2. 非精确一维线搜索（近似一维搜索）

(1) Goldstein型线搜索（1965）

选取$\lambda_k > 0$使得：$f(x^k+\lambda_k d^k)-f(x^k)\le \sigma\lambda_k \nabla f(x^k)^T d^k$,

$$f(x^k+\lambda_k d^k)-f(x^k)\ge (1-\sigma)\lambda_k \nabla f(x^k)^T d^k,$$

其中$\sigma\in(0,\frac{1}{2})$.

(2) Armijo型线搜索（1966）

选取$\lambda_k := \rho\gamma^{m_k}$使得$m_k$为满足下式的最小非负整数：

$$f(x^k+\rho\gamma^{m_k} d^k)-f(x^k)\le \sigma\rho\gamma^{m_k} \nabla f(x^k)^T d^k$$

其中$\rho>0,\ \sigma,\gamma\in(0,1)$.

(3) Wolfe型线搜索（Wolfe-Powell型线搜索）

1969 1976

选取$\lambda_k > 0$使得：$f(x^k+\lambda_k d^k)-f(x^k)\le \sigma\lambda_k \nabla f(x^k)^T d^k$,

$$\nabla f(x^k+\lambda_k d^k)^T d^k \ge \delta \nabla f(x^k)^T d^k,$$

其中$\sigma\in(0,\frac{1}{2})$, $\delta\in(\sigma,1)$.

3. 非单调一维线搜索

(1) Grippo-Lampariello-Lucidi非单调线搜索（1986）

寻找步长$\lambda_k = \rho\gamma^{h_k}$使得$h_k$是满足下式的最小非负整数：

$$f(x^k + \lambda_k d^k) \le \max_{0\le j\le m_k} f(x^{k-j}) + \delta\lambda_k \nabla f(x^k)^T d^k$$

其中，m_k是一个正整数且$m_k \le M$ (M为某一正整数），

λ_k^0是一个给定的小实数，σ，$\rho \in$（0, 1）.

注：当$f(x^k) = \max_{0\le j\le m_k} f(x^{k-j})$时，即为单调的Armijo型线搜索.

（2） Zhang-Hager（2004）非单调线搜索

寻找步长$\lambda_k > 0$,使得：

$$f(x^k + \lambda_k d^k) \le C_k + \delta\lambda_k \nabla f(x^k)^T d^k$$

$$\nabla f(x^k + \lambda_k d^k)^T d^k \ge \sigma \nabla f(x^k)^T d^k$$

其中， $0<\delta<\sigma<1$，C_k按如下方式选取：

$$C_{k+1} = (\eta_k Q_k C_k + f(x^{k+1})\ )/Q_{k+1}$$

这里，$Q_{k+1}=\eta_k Q_k + 1$， $C_0 = f(x^0), Q_0 = 1$, 且

$$\eta_k \in [\eta_{\min}, \eta_{\max}], 0 \le \eta_{\min} \le \eta_{\max} \le 1$$

注：C_k是$f(x^0), f(x^1), \cdots f(x^k)$的凸组合， 含有第$k$步迭代之前所有迭代点函数值的信息；

η_k的选取控制着“非单调性”的度， 如果$\eta_k=0$, $k \in \{0,1,2,\cdots\}$,则即为单调的$Wolfe$搜索。

（3） $Hu-Huang-Lu(2010)$非单调线搜索

寻找步长$\lambda_k>0$使得：$f(x^k+\lambda_k d^k)\le C_k+\delta\lambda_k\nabla f(x^k)^T d^k$

$$\nabla f(x^k+\lambda_k d^k)^T d^k\ge\sigma\nabla f(x^k)^T d^k$$

其中,0<δ<σ<1，C_k按如下方式选取：

$$C_{k+1}=(\eta_k\sum_{i=1}^{m_k-1}\eta_{k-i}f(x^{k-i})+f(x^k)\ )/Q_k \quad 且 Q_k=\eta_k\sum_{i=1}^{m_k-1}\eta_{k-i}+1,$$

其中$\eta_k\in[\eta_{\min},\eta_{\max}]$且$0\le\eta_{\min}\le\eta_{\max}\le 1, f(x^k)-C_k\le\dfrac{\delta l_k\left|\nabla f(x^k)^T d^k\right|}{2}$

这里，若∇f是$Lipschitz$连续的，则令$l_k:=\dfrac{(1-\sigma)\left|\nabla f(x^k)^T d^k\right|}{L\left\|d^k\right\|^2}$,

其中L为$Lipschitz$常数；否则，令$l_k:=0$

注1：C_k是目前迭代点函数值和之前若干迭代点对应函数值的凸组合，该方法综合使用了Grippo-Lampariello-Lucidi非单调线搜索和Zhang-Hager非单调线搜索的思想。

注2：一个特例：若$\sum_{i=1}^{m_k-1}\eta_{k-i}f(x^{k-i})\ge\sum_{i=1}^{m_k-1}\eta_{k-i}f(x^k)$，则选择$\eta_k\in(0,1]$;

否则，选择η_k=0。因此，每一步或者使用单调线搜索，或者使用非单调线搜索，是一种混合线搜索方法。

1.4.4 算法收敛性

1. 适定性(well-defined):如果算法的每一步是适定的，则这个算法是适定的。
2. 算法是收敛的：如果算法是有限终止的或所产生迭代点列$\{x^k\}$的每个聚点是优化问题的最优解，则称该算法是收敛的。
3. 全局收敛(globally convergent)、局部收敛(locally convergent):

   如果对于任意的初始点x^0,算法是收敛的，则称该算法是全局收敛的；

   如果只有初始点x^0充分靠近最优解时，算法是收敛的，则称该算法是局部收敛的。

4 收敛率($rate\ of\ \ convergence$)

假设算法产生无穷点列$\{x^k\}$,且算法收敛到最优解x^*,不妨设$\{x^k\}$收敛到x^*，即$\lim\limits_{k\to\infty}\left\|x^k - x^*\right\| = 0$

(1) 全局$Q-$线性收敛：

如果算法初始迭代点的选取无关于最优解的信息，且存在$\beta\in$（0, 1）使得:$\left\|x^{k+1}-x^*\right\|\leq\beta\left\|x^k - x^*\right\|, \forall k\in\{0,1,2,\cdots\}$

则称该算法是全局$Q-$线性收敛的。

（2）局部$Q-$收敛率：

如果存在$\beta > 0, \zeta\geq 0$使得：$\lim\limits_{k\to\infty}\dfrac{\left\|x^{k+1}-x^*\right\|}{\left\|x^k - x^*\right\|^\beta}=\zeta$，则称该算法是局部$Q-\beta$阶收敛的。

如果存在$\beta > 0, \zeta \geq 0$使得： $\lim\limits_{k\to\infty}\dfrac{\left\|x^{k+1}-x^{*}\right\|}{\left\|x^{k}-x^{*}\right\|^{\beta}}=\zeta$，则称该算法是局部$Q-\beta$阶收敛的。

注：特别的，

若$\beta=1, \zeta \in (0,1)$，则该算法是局部$Q-$线性收敛的；

若$\beta \in (1,2), \zeta > 0$或$\beta=1, \zeta=0$,则称该算法是局部$Q-$超线性收敛的；

若$\beta=2, \zeta > 0$，则称该算法是局部$Q-$二阶收敛的。