无约束优化线搜索方法

1

一、精确线搜索方法

优化模型 $$\min_{\boldsymbol{x}\in\mathbb{R}^n} \quad f(\boldsymbol{x})$$

其中，$f:\mathbb{R}^n \to \mathbb{R}$ 连续可微(光滑函数).

有关记号

$$\boldsymbol{g}_k = \nabla f(\boldsymbol{x}_k), \quad \boldsymbol{g}(\boldsymbol{x}) = \nabla f(\boldsymbol{x})$$

$$\boldsymbol{G}_k = \nabla^2 f(\boldsymbol{x}_k), \quad \boldsymbol{G}(\boldsymbol{x}) = \nabla^2 f(\boldsymbol{x}).$$

$$\min_{\boldsymbol{x}\in\mathbb{R}^n} \quad f(\boldsymbol{x})$$

无法一步得到最优解

求解策略：逐步调优

求解过程：

$$\boldsymbol{x}_0 \rightarrow \boldsymbol{x}_1 \rightarrow \boldsymbol{x}_2 \rightarrow \cdots\cdots$$

$$\boxed{f(\boldsymbol{x}_{k+1}) < f(\boldsymbol{x}_k)}$$

$$\rightarrow \boldsymbol{x}_k \rightarrow \boldsymbol{x}_{k+1} \rightarrow \cdots\cdots \rightarrow \boldsymbol{x}^*$$

最优解

$$\boxed{\boldsymbol{x}_k \xrightarrow{?} \boldsymbol{x}_{k+1}}$$

新迭代点

搜索方向

$$\boldsymbol{x}_{k+1} = \boldsymbol{x}_k + \alpha_k \boldsymbol{d}_k$$

当前迭代点

步长

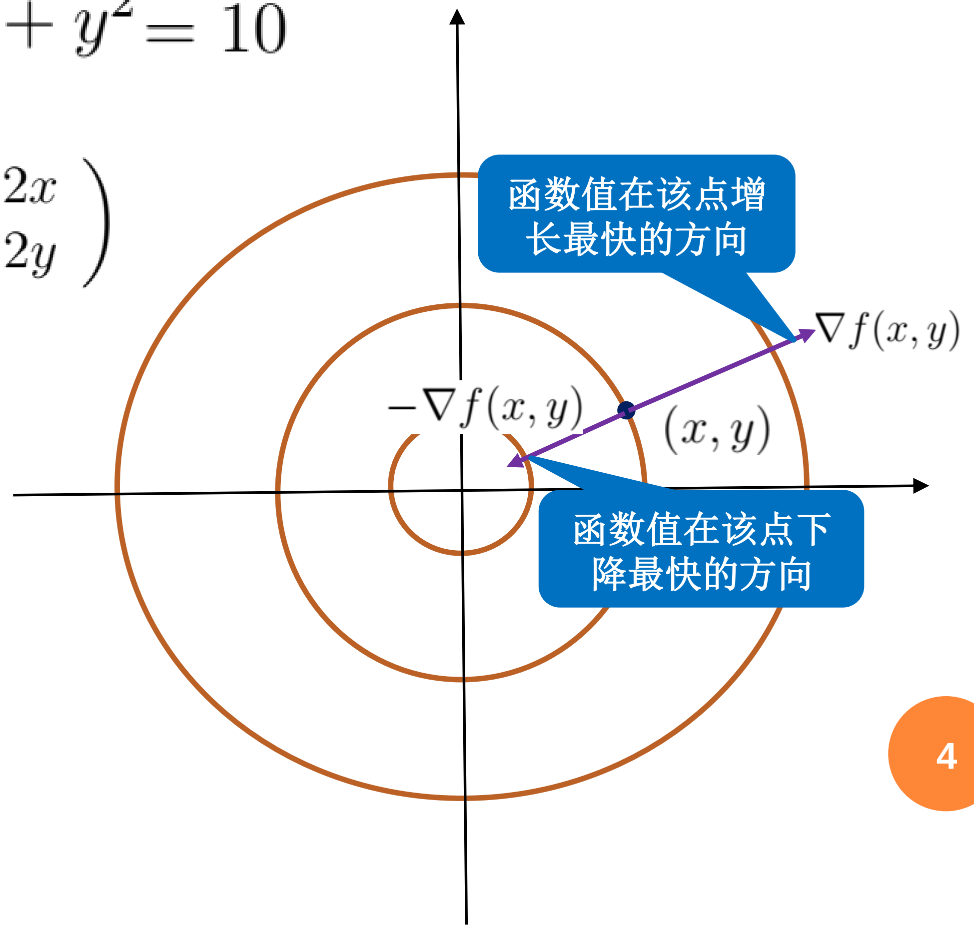
下降方向 目标函数在当前点的下降方向：存在 $\delta > 0$ 使得

$$f(\boldsymbol{x}+\alpha\boldsymbol{d}) < f(\boldsymbol{x}) \text{ 对任意 } \alpha \in (0,\delta]$$

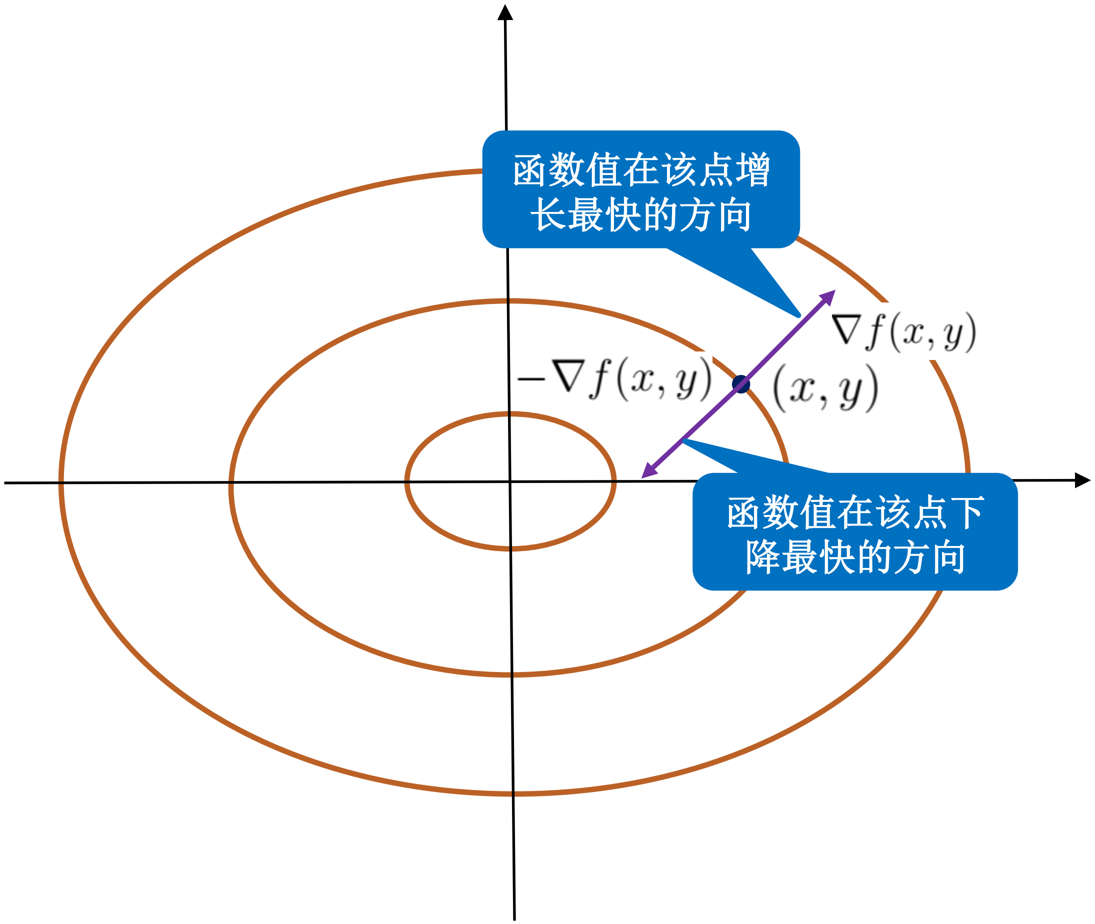
如何判定？

直观分析 $f(x,y) = x^2 + y^2 = 10$

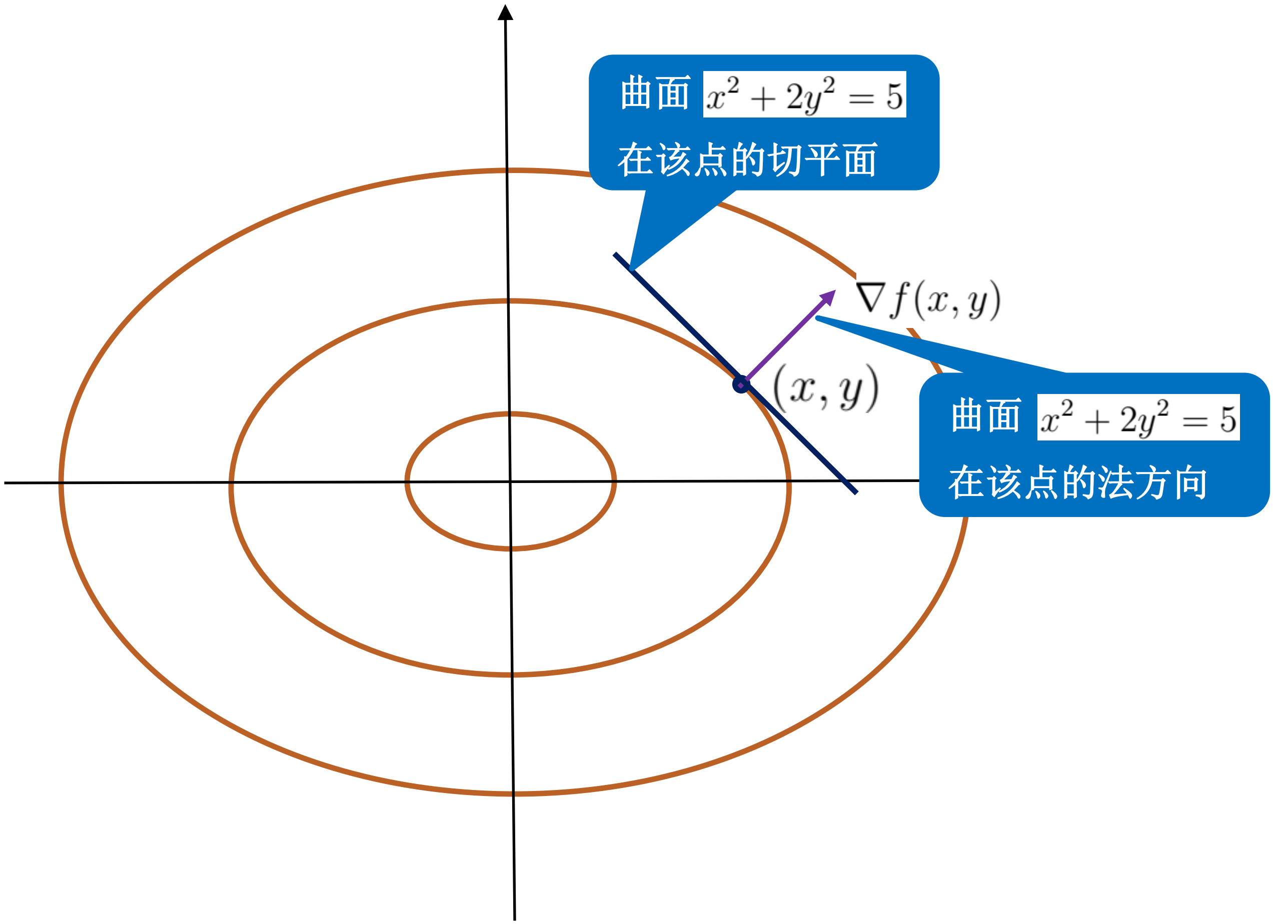
$$\nabla f(x,y) = \begin{pmatrix} \frac{\partial f(x,y)}{\partial x} \\ \frac{\partial f(x,y)}{\partial y} \end{pmatrix} = \begin{pmatrix} 2x \\ 2y \end{pmatrix}$$



$$f(x,y) = x^2 + 2y^2 = 10 \qquad \nabla f(x,y) = \begin{pmatrix} \frac{\partial f(x,y)}{\partial x} \\ \frac{\partial f(x,y)}{\partial y} \end{pmatrix} = \begin{pmatrix} 2x \\ 4y \end{pmatrix}$$

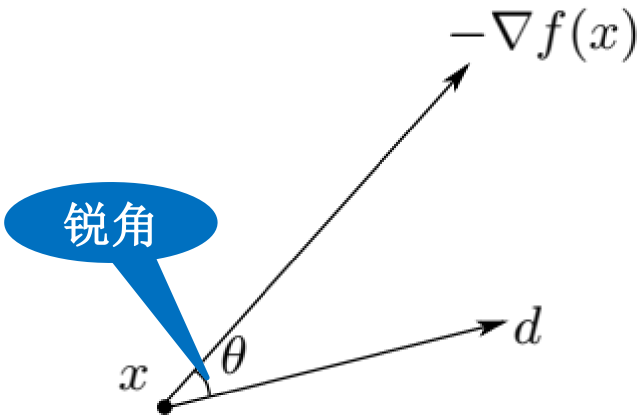


$$f(x,y)=x^2+2y^2=5 \qquad \nabla f(x,y)=\begin{pmatrix}\frac{\partial f(x,y)}{\partial x}\\ \frac{\partial f(x,y)}{\partial y}\end{pmatrix}=\begin{pmatrix}2x\\ 4y\end{pmatrix}$$



下降方向的判定： $\boldsymbol{d}^{\mathrm{T}}\nabla f(\boldsymbol{x})<0$

直观意义： $\boldsymbol{d}_k$ 与负梯度方向 $-\boldsymbol{g}_k$ 的夹角呈锐角



理论分析： $$f(\boldsymbol{x}+\alpha\boldsymbol{d})=f(\boldsymbol{x})+\alpha\nabla f(\boldsymbol{x})^{\mathrm{T}}\boldsymbol{d}+o(\alpha)$$
$$<f(\boldsymbol{x})$$

$\alpha>0$充分小

迭代过程 $\boldsymbol{x}_{k+1} = \boldsymbol{x}_k + \alpha_k \boldsymbol{d}_k$

精确线搜索 $\alpha_k = \arg\min_{\alpha \geqslant 0} f(\boldsymbol{x}_k + \alpha \boldsymbol{d}_k).$

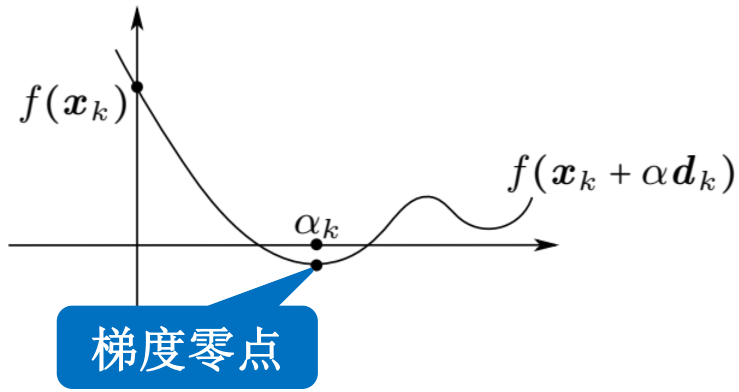
最优步长

直观意义： 沿下降方向寻求目标函数的最大下降点。

基本性质

$$f(x_k + \alpha_k d_k) \leqslant f(x_k + \alpha d_k), \quad \forall\ \alpha > 0$$

$$f'_\alpha(\boldsymbol{x}_k + \alpha_k \boldsymbol{d}_k) = \boldsymbol{d}_k^{\mathrm{T}} \nabla f(\boldsymbol{x}_k + \alpha_k \boldsymbol{d}_k) = 0$$



最优步长的计算

一元函数的极值问题, 全局最优步长难求. 通常只能求得局部最优或近似最优步长。

常用的方法：黄金分割法、多项式插值

精确线搜索下降算法

初始点　有关参数　计数器

步1、取初始点 $\boldsymbol{x}_0 \in \mathbb{R}^n$ 和参数 $\varepsilon \geqslant 0$． 令 $k = 0$

步2、 若 $\|\boldsymbol{g}_k\| \leqslant \varepsilon$， 算法终止；否则， 进入下一步.

终止规则

下降方向

步3、计算下降方向 $\boldsymbol{d}_k$ 满足 $\boldsymbol{d}_k^{\mathrm{T}} \boldsymbol{g}_k < 0$

步4、计算步长

最优步长

$$\alpha_k = \arg\min\{f(\boldsymbol{x}_k + \alpha \boldsymbol{d}_k) \mid \alpha \geqslant 0\}$$

步5、令 $\boldsymbol{x}_{k+1} = \boldsymbol{x}_k + \alpha_k \boldsymbol{d}_k,\ k = k+1$, 转步2

开始新迭代

参数选取

若算法终止规则 $\|\boldsymbol{g}_k\| \leqslant \varepsilon$ 中，$\varepsilon = 0$, 算法会产生无穷点列. 实际计算时应避免。

一般取 $\varepsilon = 10^{-4}, 10^{-6}, 10^{-8}$，$10^{-12}$，$10^{-16}$

理论分析，取 $\varepsilon = 0$, .

算法收敛性

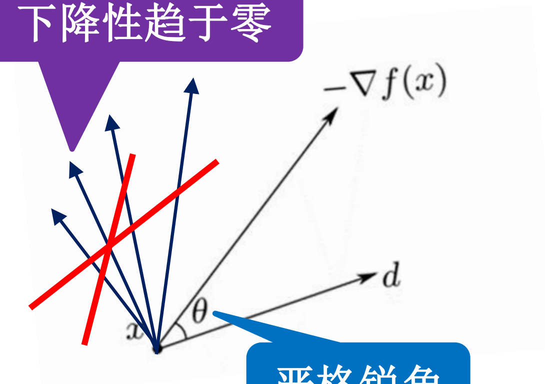
要保证算法收敛，要对搜索方向、目标函数做些假设。

定理 设目标函数 $f \in C^2$ 有下界，

搜索方向的一致下降性

下降方向 $\boldsymbol{d}_k$ 与负梯度方向 $-\boldsymbol{g}_k$ 的夹角满足 $\theta_k \leqslant \pi/2 - \theta$，$0 < \theta \leqslant \pi/2.$

下降性趋于零



严格锐角

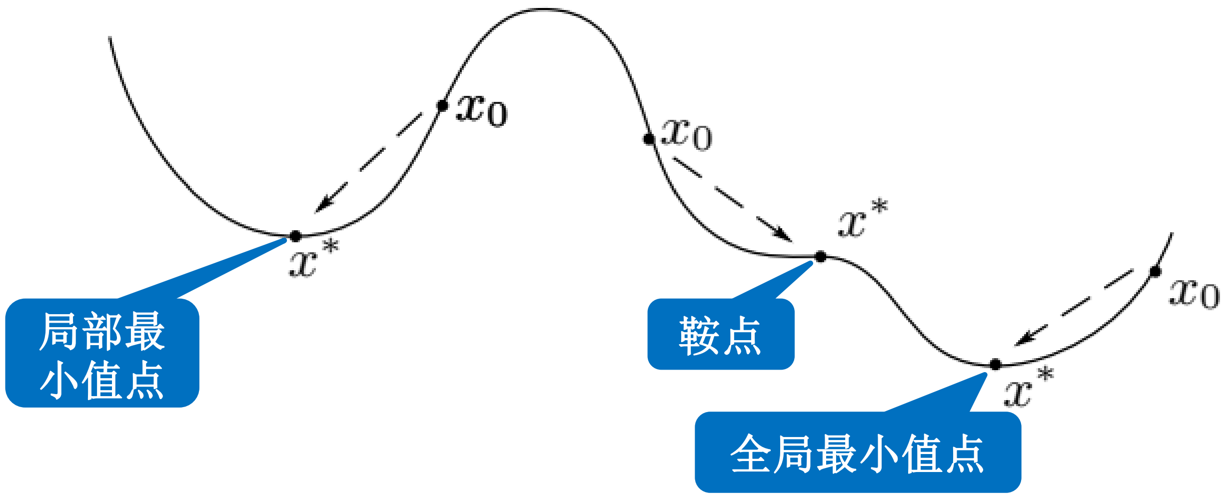
若精确线搜索方法产生无穷迭代点列，

且满足 $\|\nabla^2 f(\boldsymbol{x}_k + \alpha \boldsymbol{d}_k)\| \leqslant M, \quad \forall\ \alpha > 0$

其中，$M > 0$ 为常数，则算法收敛，即 $\boxed{\lim\limits_{k\to\infty} \boldsymbol{g}_k = \boldsymbol{0}.}$

最常见的收敛性结论

下降算法产生的聚点满足 $\lim_{k\to\infty} \boldsymbol{g}_k = \mathbf{0}$. 该聚点即便不是最小值点，也肯定不是最大值点，最差是鞍点。



为求全局最优值点，可取多个初始点进行计算，取结点中目标函数值最小的当做全局最小值点。

若目标函数梯度一致连续，则有如下结论。

定理 设目标函数 $f \in C$ 有下界，梯度函数在包含水平集的某邻域内一致连续，

一致下降性条件

$$\boxed{\text{下降方向}\ \boldsymbol{d}_k\ \text{与负梯度}\ -\boldsymbol{g}_k\ \text{的夹角满足}\ \theta_k \leqslant \pi/2-\theta,\ 0<\theta\leqslant\pi/2.}$$

若算法不有限步终止，则 $\lim\limits_{k\to\infty} \boldsymbol{g}_k = \mathbf{0}.$

水平集 $\mathcal{L}(\boldsymbol{x}_0)=\{\boldsymbol{x}\in\mathbb{R}^n \mid f(\boldsymbol{x})\leqslant f(\boldsymbol{x}_0)\}$

一致连续 对任意的 $\varepsilon>0$，存在 $\delta>0$，使对任意满足 $\|\boldsymbol{x}-\boldsymbol{y}\|\leqslant\delta$ 的 $\boldsymbol{x},\boldsymbol{y}\in\mathbb{R}^n$，均有 $|f(\boldsymbol{x})-f(\boldsymbol{y})|\leqslant\varepsilon$

较弱假设条件下的收敛性

定理 设目标函数 $f \in C$. 若存在收敛子列，使得

$$\lim_{j\to\infty} \boldsymbol{x}_{k_j} = \boldsymbol{x}^*, \quad \lim_{j\to\infty} \boldsymbol{d}_{k_j} = \boldsymbol{d}^*$$

，则 $\boldsymbol{g}(\boldsymbol{x}^*)^{\mathrm{T}}\boldsymbol{d}^* = 0.$

进一步，若目标函数 $f \in C^2$, 则 $(\boldsymbol{d}^*)^{\mathrm{T}}\nabla^2 f(\boldsymbol{x}^*)\boldsymbol{d}^* \geqslant 0.$

注：未对搜索方向和目标函数梯度做要求，因而结论相对较弱。

收敛速度估计

定理 设搜索方向与负梯度夹角满足前面结论中的条件，

即下降方向 $\boldsymbol{d}_k$ 与负梯度 $-\boldsymbol{g}_k$ 的夹角满足 $\theta_k \leqslant \pi/2-\theta$

其中 $0<\theta\leqslant\pi/2.$

算法产生的迭代点列收敛到目标函数的最小值点 $\boldsymbol{x}^*$；

目标函数在 $\boldsymbol{x}^*$ 点严格凸。则迭代点列R-线性收敛。

注：线性收敛是一个比较慢的收敛速度。

但也要求比较强的条件。

最优步长小结

$$\alpha_k = \arg\min_{\alpha \geq 0} f(x_k + \alpha d_k)$$

- 目标函数沿搜索方向下降量达到最大。

  理想化策略，计算量大，很难取到。

- 便于算法理论分析（如二次终止性、收敛速率等）

- 对全局最优解不起关键作用

二、非精确线搜索方法

最优步长的初衷是在每一迭代步使目标函数下降量达到最大. 但计算量太大，而且不好求.

最优化问题关注全局最优值点。集中于某个方向上的线搜索似乎没有必要。

常见的非精确步长规则

- Armijo步长
- Wolfe步长

1. Armijo步长规则

设 $\boldsymbol{d}_k$为$\boldsymbol{x}_k$点的下降方向，即 $\boldsymbol{d}_k^{\mathrm{T}}\boldsymbol{g}_k < 0$

对 $\sigma \in (0,1)$，只要$\alpha > 0$充分小

$$f(x_k + \alpha d_k) = f(x_k) + \boxed{\alpha g_k^{\mathrm{T}} d_k} + o(\alpha)$$

$$< 0$$

步长选取：先取一个较大的步长，看目标函数是否有满意的下降量。否则依次按比例压缩，直至满足要求。又称进退试探法。

Armijo步长规则

取常数 $\beta > 0,\ \sigma,\ \gamma \in (0,1)$。步长 $\alpha_k = \beta\gamma^{m_k}$,

其中 m_k 为 0, 1, 2, 3, …中满足下式的最小值

$$f(x_k + \beta\gamma^m d_k) \leqslant f_k + \boxed{\sigma\beta\gamma^m g_k^{\mathrm{T}} d_k.}$$

下降量

若 $\alpha_k < \beta$，则

$$f(\boldsymbol{x}_k + \beta\gamma^{\boxed{m_k}}\boldsymbol{d}_k) \leqslant f(\boldsymbol{x}_k) + \sigma\beta\gamma^{m_k}\boldsymbol{g}_k^{\mathrm{T}}\boldsymbol{d}_k,$$

$$f(\boldsymbol{x}_k + \beta\gamma^{\boxed{m_k-1}}\boldsymbol{d}_k) > f(\boldsymbol{x}_k) + \sigma\beta\gamma^{m_k-1}\boldsymbol{g}_k^{\mathrm{T}}\boldsymbol{d}_k.$$

2、Wolfe步长

常数 $0<\sigma_1<\sigma_2<1$. 步长 α_k 同时满足

下降性条件

$$f(\boldsymbol{x}_k+\alpha\boldsymbol{d}_k)\leqslant f(\boldsymbol{x}_k)+\sigma_1\alpha\boldsymbol{g}_k^{\mathrm{T}}\boldsymbol{d}_k,$$

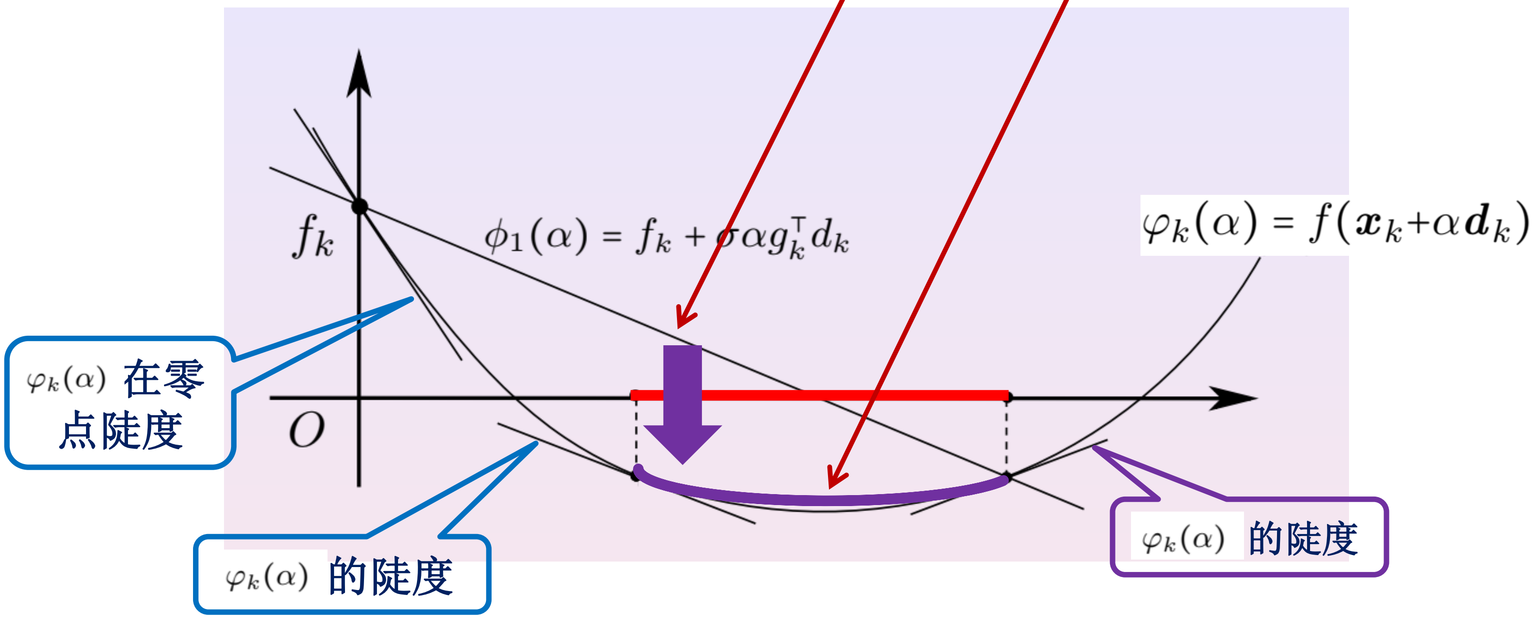
$$\nabla f(\boldsymbol{x}_k+\alpha\boldsymbol{d}_k)^{\mathrm{T}}\boldsymbol{d}_k\geqslant\sigma_2\boldsymbol{g}_k^{\mathrm{T}}\boldsymbol{d}_k. \quad \Longleftrightarrow \quad \boldsymbol{g}_{k+1}^{\mathrm{T}}\boldsymbol{d}_k\geqslant\sigma_2\boldsymbol{g}_k^{\mathrm{T}}\boldsymbol{d}_k.$$

$\varphi_k(\alpha)=f(\boldsymbol{x}_k+\alpha\boldsymbol{d}_k)$ 的陡度比 $\alpha=0$ 点有所减缓

从而远离当前迭代点

$$f(\boldsymbol{x}_k+\alpha\boldsymbol{d}_k)\leqslant f(\boldsymbol{x}_k)+\sigma_1\alpha\boldsymbol{g}_k^{\mathrm{T}}\boldsymbol{d}_k,$$
$$\nabla f(\boldsymbol{x}_k+\alpha\boldsymbol{d}_k)^{\mathrm{T}}\boldsymbol{d}_k\geqslant\sigma_2\boldsymbol{g}_k^{\mathrm{T}}\boldsymbol{d}_k.$$

步长的存在性



加强版：强Wolfe步长规则

$$\nabla f(x_k + \alpha d_k)^T d_k \geq \sigma_2 g_k^T d_k$$

$$\left|\nabla f(x_k + \alpha d_k)^T d_k\right| \leq \sigma_2 \left|g_k^T d_k\right|$$

- 若 $\sigma_2=0$，即 $g_{k+1}^T d_k=0$，可视作最优步长
- Wolfe步长只考虑 $\boldsymbol{g}_{k+1}^{\mathrm{T}} \boldsymbol{d}_k < 0$ 情形，可能产生差步长；强Wolfe步长可避免。

几种非精确步长的异同

相同点

均要求目标函数有满意的下降量且控制步长不能太小。

相异点

Armijo步长最为常见；Wolfe步长包含最优步长，特别适用于共轭梯度与拟牛顿方法；Armijo步长通过进退试探技术；Wolfe步长借助多项式插值等技术。

非精确线搜索算法

步1. 取初始点 $\boldsymbol{x}_0 \in \mathbb{R}^n$ 和参数 $\varepsilon \geqslant 0$. 令 $k = 0$

步2. 若 $\|\boldsymbol{g}_k\| \leqslant \varepsilon$, 算法终止；否则，进入下一步.

步3. 计算下降方向 $\boldsymbol{d}_k$，使满足 $\boldsymbol{d}_k^{\mathrm{T}} \boldsymbol{g}_k < 0$

步4. 利用非精确步长规则计算步长 α_k。

步5. 令 $\boldsymbol{x}_{k+1} = \boldsymbol{x}_k + \alpha_k \boldsymbol{d}_k$, $k = k+1$, 转步2.

算法收敛性

定理　设目标函数连续可微。Armijo步长规则下，算法产生迭代点列的任一聚点为目标函数的稳定点.

证明： 以 $\boldsymbol{d}_k = -\boldsymbol{g}_k$ 为例证明。

函数值列 $\{f(\boldsymbol{x}_k)\}$ 单调下降，迭代点列有聚点，函数值列有界。

下降量趋于零

$$\boxed{\lim_{k\to\infty}\big(f(\boldsymbol{x}_k) - f(\boldsymbol{x}_{k+1})\big) = 0}.$$

反设结论不成立，即

$$\lim_{\substack{k\in K\\k\to\infty}} \boldsymbol{d}_k^{\mathrm{T}}\nabla f(\boldsymbol{x}_k)=\lim_{\substack{k\in K\\k\to\infty}} -\|\nabla f(\boldsymbol{x}_k)\|^2<-\varepsilon_0$$

由线索搜过程

$$f(\boldsymbol{x}_k+\beta\gamma^{m_k}\boldsymbol{d}_k)\leqslant f(\boldsymbol{x}_k)+\sigma\beta\gamma^{m_k}\boldsymbol{g}_k^{\mathrm{T}}\boldsymbol{d}_k,$$

$$\Downarrow$$

$$\lim_{\substack{k\in \mathcal{K}\\k\to\infty}} \alpha_k=0.$$

$$f(\boldsymbol{x}_k+(\alpha_k/\gamma)\boldsymbol{d}_k)-f(\boldsymbol{x}_k)>\sigma(\alpha_k/\gamma)\nabla f(\boldsymbol{x}_k)^{\mathrm{T}}\boldsymbol{d}_k.$$

$$f(\boldsymbol{x}_k + (\alpha_k/\gamma)\boldsymbol{d}_k) - f(\boldsymbol{x}_k) > \sigma(\alpha_k/\gamma)\nabla f(\boldsymbol{x}_k)^{\mathrm{T}}\boldsymbol{d}_k.$$

$\bar{\alpha}_k \triangleq \alpha_k/\gamma$

存在 $\hat{\alpha}_k \in [0, \bar{\alpha}_k]$

$$\nabla f(\boldsymbol{x}_k + \hat{\alpha}_k \boldsymbol{d}_k)^{\mathrm{T}}\boldsymbol{d}_k > \sigma\nabla f(\boldsymbol{x}_k)^{\mathrm{T}}\boldsymbol{d}_k, \qquad \forall\ k \in \mathcal{K},\ k \geqslant K.$$

取极限

$$\lim_{\substack{k\in K\\ k\to\infty}} \nabla f(\boldsymbol{x}_k)^{\mathrm{T}}\boldsymbol{d}_k \geqslant \lim_{\substack{k\in K\\ k\to\infty}} \sigma\nabla f(\boldsymbol{x}_k)^{\mathrm{T}}\boldsymbol{d}_k$$

$$(1-\sigma)\lim_{\substack{k\in K\\ k\to\infty}} \nabla f(\boldsymbol{x}_k)^{\mathrm{T}}\boldsymbol{d}_k \geqslant 0 \quad \xleftrightarrow{\text{矛盾}} \quad \lim_{\substack{k\in K\\ k\to\infty}} \boldsymbol{d}_k^{\mathrm{T}}\nabla f(\boldsymbol{x}_k) < -\varepsilon_0.$$

定理（Wolfe步长）设目标函数连续可微有下界，梯度在水平集上Lipschitz连续.则Wolfe步长规则下，算法产生的点列满足

$$\sum_{k=1}^{\infty} \|\boldsymbol{g}_k\|^2 \cos^2(\boldsymbol{d}_k, -\boldsymbol{g}_k) < \infty.$$

定理（Wolfe步长）设目标函数连续可微有下界，梯度在水平集上一致连续. 则Wolfe线搜索下降算法产生的点列满足

$$\lim_{k\to\infty} \frac{\boldsymbol{g}_k^{\mathrm{T}} \boldsymbol{d}_k}{\|\boldsymbol{d}_k\|} = 0.$$

信赖域方法

一、信赖域方法

优化模型 $\min\limits_{\boldsymbol{x}\in\mathbb{R}^n} \quad f(\boldsymbol{x})$

线搜索方法 $\boldsymbol{x}_k \xrightarrow{?} \boldsymbol{x}_{k+1} \qquad f(\boldsymbol{x}_{k+1}) < f(\boldsymbol{x}_k)$

$$\boldsymbol{x}_{k+1} = \boldsymbol{x}_k + \alpha_k \boldsymbol{d}_k$$

信赖域方法： $\boldsymbol{x}_k \longrightarrow \boldsymbol{x}_{k+1} \qquad f(\boldsymbol{x}_{k+1}) < f(\boldsymbol{x}_k)$

算法思想： 构造目标函数的二阶近似, 将其在当前点邻域内的最小值点作为新的迭代点.

思想：依据二次模型与目标函数的近似度调节信赖域半径：若新迭代点不能使目标函数有充分下降，就缩小信赖域半径；否则，扩大信赖域半径。

关键技术：信赖域半径的调整。

$$\min_{\boldsymbol{x}\in\mathbb{R}^n} \quad f(\boldsymbol{x})$$

构造 $f(\boldsymbol{x}_k+\boldsymbol{d})$的二阶近似

$$m_k(\boldsymbol{d})=f(\boldsymbol{x}_k)+\boldsymbol{g}_k^{\mathrm{T}}\boldsymbol{d}+\frac{1}{2}\boldsymbol{d}^{\mathrm{T}}\boldsymbol{B}_k\boldsymbol{d}$$

其中 $\boldsymbol{B}_k=\nabla^2 f(\boldsymbol{x}_k)$ 或 $\boldsymbol{B}_k\approx\nabla^2 f(\boldsymbol{x}_k)$.

用二次模型的最小值点 $\boldsymbol{d}_k$ 当做 $f(\boldsymbol{x}_k+\boldsymbol{d})$ 关于 $\boldsymbol{d}$ 的最小值点。

$f(\boldsymbol{x}_k+\boldsymbol{d})$ 与 $m_k(\boldsymbol{d})$ 的近似度高低取决于 $\boldsymbol{d}$ 的模大小

信赖域子问题 $\min\{m_k(\boldsymbol{d}) \mid \boldsymbol{d} \in \mathbb{R}^n,\ \|\boldsymbol{d}\| \leqslant \Delta_k\}$,

其中，$\Delta_k > 0$ 为信赖域半径

设 $\boldsymbol{d}_k$ 为其最优解。则目标函数与二次模型的近似度为

$$f(\boldsymbol{x}_k + \boldsymbol{d}_k) - m_k(\boldsymbol{d}_k) = O(\|\boldsymbol{d}_k\|^2)$$

若 $\boldsymbol{B}_k = \nabla^2 f(\boldsymbol{x}_k)$，则

$$f(\boldsymbol{x}_k + \boldsymbol{d}_k) - m_k(\boldsymbol{d}_k) = o(\|\boldsymbol{d}_k\|^2)$$

信赖域半径调整

调整依据： $m_k(\boldsymbol{d}_k)$ 与 $f(\boldsymbol{x}_k+\boldsymbol{d}_k)$ 的近似度

模型下降量

预下降量 $\mathrm{Pred}_k = m_k(\boldsymbol{0}) - m_k(\boldsymbol{d}_k),$

下降比 $r_k = \dfrac{\mathrm{Ared}_k}{\mathrm{Pred}_k}$

实下降量 $\mathrm{Ared}_k = f(\boldsymbol{x}_k) - f(\boldsymbol{x}_k + \boldsymbol{d}_k)$

- 若 $r_k < 0$, $\mathrm{Ared}_k < 0$, $\boldsymbol{x}_k + \boldsymbol{d}_k$ 不能作为新迭代点, 缩小信赖域半径，重新计算 $\boldsymbol{d}_k$ ；
- 若 $r_k > 0$ 且靠近1，近似度较好。令 $\boldsymbol{x}_{k+1} = \boldsymbol{x}_k + \boldsymbol{d}_k$ 。扩大信赖域半径；
- 其他情况，$\boldsymbol{x}_{k+1} = \boldsymbol{x}_k + \boldsymbol{d}_k$ ，不调整信赖域半径.

信赖域算法

步1、取 $\hat{\Delta} > 0$，$\boldsymbol{x}_0 \in \mathbb{R}^n$，$\Delta_0 \in (0, \hat{\Delta}]$，$\eta \in [0, \frac{1}{4})$，$\varepsilon \geqslant 0$. 令 $k = 0$

步2、若 $\|\boldsymbol{g}_k\| \leqslant \varepsilon$，算法终止；否则，进入下一步.

步3、解子问题 $\boldsymbol{d}_k = \arg\min\{m_k(\boldsymbol{d}) \mid \boldsymbol{d} \in \mathbb{R}^n,\ \|\boldsymbol{d}\| \leqslant \Delta_k\}$

计算 $r_k = \dfrac{\mathrm{Ared}_k}{\mathrm{Pred}_k}$

若 $r_k < \dfrac{1}{4}$，令 $\Delta_{k+1} = \dfrac{1}{4}\Delta_k$；若 $r_k > \dfrac{3}{4}$ 且 $\|\boldsymbol{d}_k\| = \Delta_k$，

令 $\Delta_{k+1} = \min\{2\Delta_k, \hat{\Delta}\}$；否则，令 $\Delta_{k+1} = \Delta_k$.

步4、若 $r_k > \eta$，令 $\boldsymbol{x}_{k+1} = \boldsymbol{x}_k + \boldsymbol{d}_k$；

否则，令 $\boldsymbol{x}_{k+1} = \boldsymbol{x}_k$. $k = k+1$，转步2.

二、子问题求解

子问题

$$\begin{aligned} &\min\ m_k(\boldsymbol{d}) = f(\boldsymbol{x}_k) + \boldsymbol{g}_k^{\mathrm{T}}\boldsymbol{d} + \frac{1}{2}\boldsymbol{d}^{\mathrm{T}}\boldsymbol{B}_k\boldsymbol{d} \\ &\text{s.t. } \|d\| \leqslant \Delta_k \end{aligned}$$

一个启发式方法（负梯度法）

将 $\boldsymbol{d}$ 限制在负梯度方向上, 即令 $\boldsymbol{d} = \tau\boldsymbol{d}_k^s$, 其中 $\boldsymbol{d}_k^s = -\boldsymbol{g}_k$

单元函数极值问题

$$\begin{aligned} &\min\ m_k(\tau\boldsymbol{d}_k^s) = f(\boldsymbol{x}_k) - \tau\|\boldsymbol{g}_k\|^2 + \frac{1}{2}\tau^2\boldsymbol{g}_k^{\mathrm{T}}\boldsymbol{B}_k\boldsymbol{g}_k \\ &\text{s.t. } \tau \geqslant 0, \quad \|\tau\boldsymbol{d}_k^s\| \leqslant \Delta_k \end{aligned}$$

$$\tau_k = \begin{cases} \dfrac{\Delta_k}{\|\boldsymbol{g}_k\|}, & \text{若 } \boldsymbol{g}_k^{\mathrm{T}}\boldsymbol{B}_k\boldsymbol{g}_k \leqslant 0, \\ \min\left\{\dfrac{\|\boldsymbol{g}_k\|^2}{\boldsymbol{g}_k^{\mathrm{T}}\boldsymbol{B}_k\boldsymbol{g}_k}, \dfrac{\Delta_k}{\|\boldsymbol{g}_k\|}\right\}, & \text{否则.} \end{cases}$$

子问题近似解：

Cauchy点

$$\Rightarrow \boldsymbol{d}_k^C = \tau_k\boldsymbol{d}_k^s$$

三种求解算法

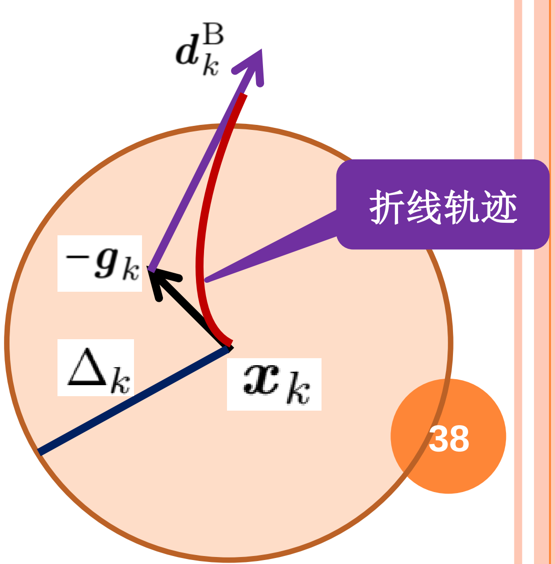
1、折线方法

信赖域半径很小时, 目标函数线性近似。最优解 $d_k^c = \tau_k d_k^s$

信赖域半径较大时, 目标函数需二阶近似，最优解

$$\boldsymbol{d}_k^{\mathrm{B}} \triangleq -\boldsymbol{B}_k^{-1}\boldsymbol{g}_k$$

介于二者之间时, 最优解形成一条由 d_k^c 到 d_k^{B} 的曲线(折线)——dogleg方法(Powell, 1970).



信赖域子问题

$$\min m_k(\boldsymbol{d}) = f(\boldsymbol{x}_k) + \boldsymbol{g}_k^{\mathrm{T}}\boldsymbol{d} + \frac{1}{2}\boldsymbol{d}^{\mathrm{T}}\boldsymbol{B}_k\boldsymbol{d}$$
$$\text{s.t.} \ \|d\| \leqslant \Delta_k$$

单元函数极值问题

令最优解为

$$\boldsymbol{d}_k(\tau) = \begin{cases} \boxed{\tau \boldsymbol{d}_k^u}, & 若\tau \in [0,1] \\ \boxed{\boldsymbol{d}_k^u + (\tau - 1)(\boldsymbol{d}_k^{\mathrm{B}} - \boldsymbol{d}_k^u)}, & 若\tau \in [1,2] \end{cases}$$

参数值待定

其中 $\boldsymbol{d}_k^u = -\dfrac{\|\boldsymbol{g}_k\|^2}{\boldsymbol{g}_k^{\mathrm{T}}\boldsymbol{B}_k\boldsymbol{g}_k}\boldsymbol{g}_k \qquad \boldsymbol{d}_k^{\mathrm{B}} \triangleq -\boldsymbol{B}_k^{-1}\boldsymbol{g}_k$

2、二维子空间方法

将信赖域子问题

$$\min m_k(\boldsymbol{d}) = f(\boldsymbol{x}_k) + \boldsymbol{g}_k^{\mathrm{T}}\boldsymbol{d} + \frac{1}{2}\boldsymbol{d}^{\mathrm{T}}\boldsymbol{B}_k\boldsymbol{d}$$
$$\text{s.t. } \|d\| \leqslant \Delta_k$$

可行域限制到子空间 $\operatorname{span}[\boldsymbol{g}_k, \boldsymbol{B}_k^{-1}\boldsymbol{g}_k]$.

$$\min m_k(\boldsymbol{d}) = f(\boldsymbol{x}_k) + \boldsymbol{g}_k^{\mathrm{T}}\boldsymbol{d} + \frac{1}{2}\boldsymbol{d}^{\mathrm{T}}\boldsymbol{B}_k\boldsymbol{d}$$
$$\text{s.t. } \|\boldsymbol{d}\| \leqslant \Delta_k$$
$$\boldsymbol{d} \in \operatorname{span}[\boldsymbol{g}_k, \boldsymbol{B}_k^{-1}\boldsymbol{g}_k]$$

3、精确解方法

利用最优性理论对信赖域子问题

$$
\begin{aligned}
&\min\ m_k(\boldsymbol{d}) = f(\boldsymbol{x}_k) + \boldsymbol{g}_k^{\mathrm{T}}\boldsymbol{d} + \frac{1}{2}\boldsymbol{d}^{\mathrm{T}}\boldsymbol{B}_k\boldsymbol{d} \\
&\text{s.t. } \|d\| \leqslant \Delta_k
\end{aligned}
$$

的最优解进行刻画，得最优解满足

$$
\begin{cases}
\|\boldsymbol{d}_k\| \leqslant \Delta_k, \\
\left(\boldsymbol{B}_k + \mu_k \boldsymbol{I}\right)\boldsymbol{d}_k = -\boldsymbol{g}_k, \\
\mu_k\left(\Delta_k - \|\boldsymbol{d}_k\|\right) = 0, \\
\left(\boldsymbol{B}_k + \mu_k \boldsymbol{I}\right)\text{半正定}.
\end{cases}
$$

注：只是对最优解进行了刻画，并未给出最优解。

子问题算法小结

- 负梯度法、折线方法、二维子空间方法给出的都是近似解，精确解方法难于计算；
- 折线方法法、二维子空间方法给出的解都优于Cauchy点。

三、方法收敛性

引理 (子问题下降量估计)信赖域子问题

$$\min\ m_k(\boldsymbol{d}) = f(\boldsymbol{x}_k) + \boldsymbol{g}_k^{\mathrm{T}}\boldsymbol{d} + \frac{1}{2}\boldsymbol{d}^{\mathrm{T}}\boldsymbol{B}_k\boldsymbol{d}$$

$$\text{s.t. } \|d\| \leqslant \Delta_k$$

在负梯度方向上的最优解(Cauchy点) $\boldsymbol{d}_k^c = \tau_k \boldsymbol{d}_k^s$

$$\tau_k = \begin{cases} \dfrac{\Delta_k}{\|\boldsymbol{g}_k\|}, & \text{若 } \boldsymbol{g}_k^{\mathrm{T}}\boldsymbol{B}_k\boldsymbol{g}_k \leqslant 0, \\ \min\left\{\dfrac{\|\boldsymbol{g}_k\|^2}{\boldsymbol{g}_k^{\mathrm{T}}\boldsymbol{B}_k\boldsymbol{g}_k}, \dfrac{\Delta_k}{\|\boldsymbol{g}_k\|}\right\}, & \text{否则.} \end{cases}$$

满足 $m_k(\boldsymbol{d}_k^c) - m_k(\boldsymbol{0}) \leqslant -\dfrac{1}{2}\|\boldsymbol{g}_k\| \min\{\Delta_k, \dfrac{\|\boldsymbol{g}_k\|}{\|\boldsymbol{B}_k\|}\}.$

注：折线方法、二维子空间方法给出的最优解均满足该性质！

算法收敛性

定理 设目标函数在水平集 $\mathcal{L}(\boldsymbol{x}_0)$ 上连续可微有下界

信赖域算法满足 $\eta = 0$，存在 $\beta > 0$，使对任意的 k，

$\|\boldsymbol{B}_k\| \leqslant \beta$，信赖域子问题的解 $\boldsymbol{d}_k$ 满足

$$m_k(\boldsymbol{d}_k) - m_k(\boldsymbol{0}) \leqslant -\frac{1}{2}\|\boldsymbol{g}_k\| \min\{\Delta_k, \frac{\|\boldsymbol{g}_k\|}{\|\boldsymbol{B}_k\|}\}$$

则算法产生的点列满足

$$\liminf_{k\to\infty} \|\boldsymbol{g}_k\| = 0.$$

定理　设目标函数梯度Lipschitz连续，在水平集上有下界；信赖域算法满足 $\eta \in (0, \frac{1}{4})$，存在 $\beta > 0$ 使对任意的 k，$\|\boldsymbol{B}_k\| \leqslant \beta$. 信赖域子问题的解 $\boldsymbol{d}_k$ 满足

$$m_k(\boldsymbol{d}_k) - m_k(\boldsymbol{0}) \leqslant -\frac{1}{2}\|\boldsymbol{g}_k\| \min\{\Delta_k, \frac{\|\boldsymbol{g}_k\|}{\|\boldsymbol{B}_k\|}\}$$

则算法产生的点列满足

$$\lim_{k\to\infty} \boldsymbol{g}_k = \boldsymbol{0}.$$