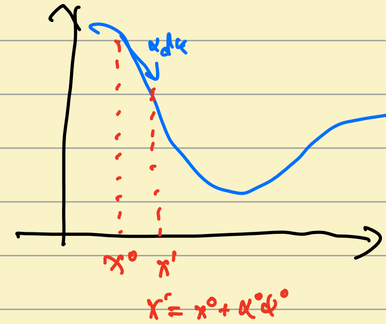
$$\min_{x} f(x) \qquad x\in\mathbb{R}^n$$



初始点：x^0 ① 方向 ② 步长

$$x^{k+1} = x^k + \alpha^k d^k$$

① 最优的搜索方向：

下降方向：$f(x^{k+1}) \le f(x^k)$ ⇒ 满足 $(d^k)^T \nabla f(x^k) < 0$

最速下降方向：$d^k = -\nabla f(x^k)$ ⇒ 最速下降法（梯度下降法）

$$x^{k+1} = x^k - \alpha \nabla f(x^k)$$

$\alpha \in (0, \frac{1}{L})$ → 合理步长

$\alpha_k = \arg\min\limits_{\alpha\ge 0} f(x^k + \alpha d^k)$ → 最优步长

① 计算量太大，不好找

② 关注全局最优，集中于某方向上的最优无必要

非精确线搜索方法

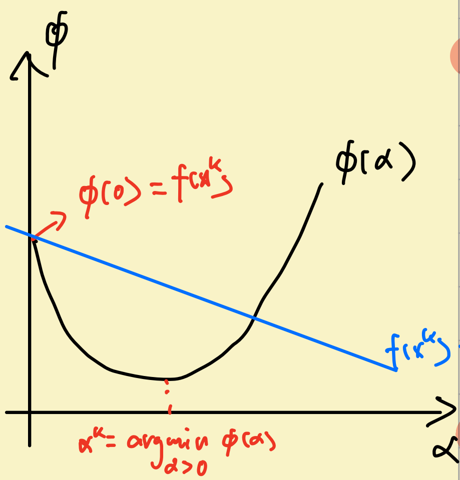
Armijo 准则

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

Defn：设 d^k 是 x^k 处的下降方向，若

$$f(x^k + \alpha d^k) \le f(x^k) + c_1 \alpha \langle \nabla f(x^k), d^k \rangle$$

则称 α 满足 Armijo 准则。$c_1 \in (0,1)$ 常数。 $c_1 = 10^{-3}$



$f(x^k) + c_1\alpha\langle \nabla f(x^k), d^k\rangle$

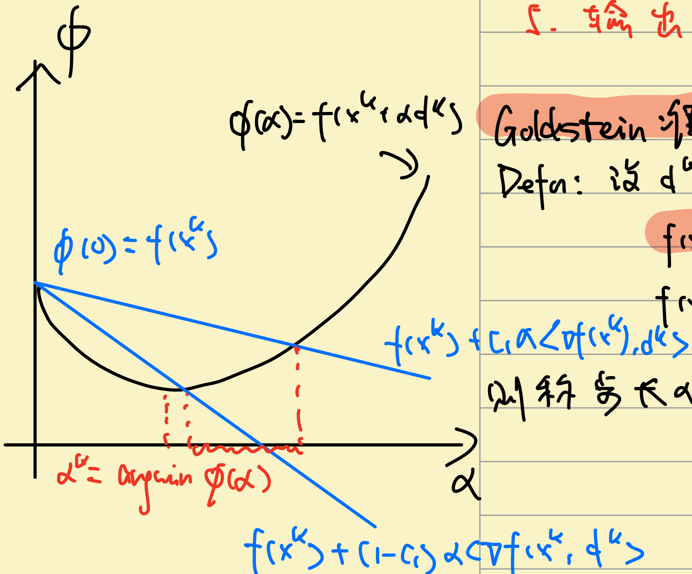
$x^{k+1} = x^k + \alpha d^k$ 令 $\phi(\alpha) = f(x^k + \alpha d^k)$

缺点：$\alpha = 0$ 显然满足条件，但无意义。采用过小的步长效率很低

回退法：从大到小搜步长，以确保α尽可能大

算法：回退法　　　　　　　　　　　　　　　　$\gamma=0.1$

1. 选择初始步长 α^0，参数 $\gamma, c\in(0,1)$　　　　$\alpha^0=1$
2. while $f(x^k+\alpha d^k) > f(x^k) + c\alpha\langle\nabla f(x^k), d^k\rangle$ do
3. 　$\alpha^{k+1} = \gamma\cdot\alpha^k$
4. End while
5. 输出 α^{k+1}



Goldstein 准则　引入下界防止步长过小

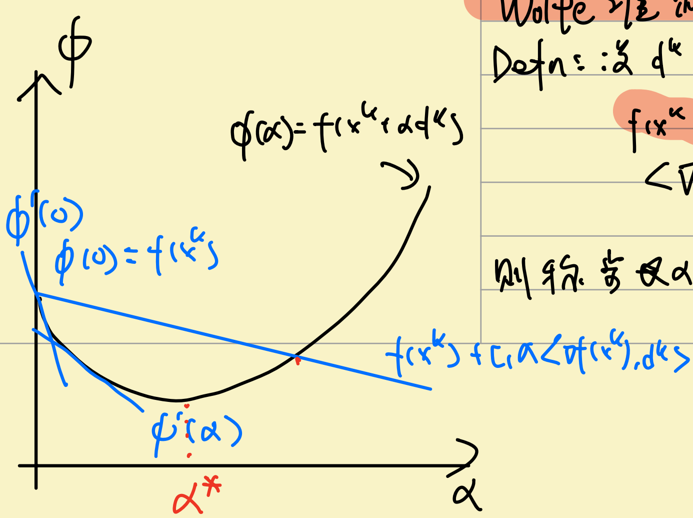
Defn：设 d^k 是 x^k 处的下降方向，若

$$f(x^k+\alpha d^k) \le f(x^k) + c_1\alpha\langle\nabla f(x^k), d^k\rangle \qquad c=10^{-3}$$

$$f(x^k+\alpha d^k) \ge f(x^k) + (1-c_1)\alpha\langle\nabla f(x^k), d^k\rangle$$

则称步长α满足Goldstein准则，其中$c_1\in(0,1)$常数.

缺点：下界可能过大，以致于跳开了最优的函数值



Wolfe 准则

Defn：设 d^k 是 x^k 处的下降方向，若

$$f(x^k+\alpha d^k) \le f(x^k) + c_1\alpha\langle\nabla f(x^k), d^k\rangle$$

$$\langle\nabla f(x^k+\alpha d^k), d^k\rangle \ge c_2\langle\nabla f(x^k), d^k\rangle$$

则称步长α满足Wolfe准则，其中 $0<c_1<c_2<1$

$\phi'(\alpha) = \langle \nabla f(x^k + \alpha d^k), d^k \rangle$

$\phi'(0) = \langle \nabla f(x^k), d^k \rangle$

$\phi'(\alpha^*) = 0 \rightarrow$ 最优步长的斜率

$0 > c_2 \langle \nabla f(x^k), d^k \rangle$

优点：一定包含最优步长

输入步长不会太小

缺点：计算量较高

收敛性分析

Thm (Zoutendijk) 设目标函数 f 有下界，连续可微，且 L-梯度 Lipschitz 连续，且在迭代中 Wolfe 准则满足，那么

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

$\cos\theta_k$ 为负梯度 $-\nabla f(x)$ 和下降方向 d^k 夹角的余弦。

$$\cos\theta_k = \frac{\langle -\nabla f(x^k), d^k \rangle}{\|\nabla f(x^k)\| \|d^k\|}$$

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

想得到

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

Wolfe condition:

$$f(x^k + \alpha d^k) \le f(x^k) + c_1 \alpha \langle \nabla f(x^k), d^k \rangle$$
$$\langle \nabla f(x^k + \alpha d^k), d^k \rangle \ge c_2 \langle \nabla f(x^k), d^k \rangle$$

Proof: $\langle \nabla f(x^k + \alpha^k d^k), d^k\rangle - \langle \nabla f(x^k), d^k\rangle \geq c_2 \langle \nabla f(x^k), d^k\rangle - \langle \nabla f(x^k), d^k\rangle$

$$\Rightarrow \langle \nabla f(x^k + \alpha^k d^k) - \nabla f(x^k), d^k\rangle \geq (c_2 - 1)\langle \nabla f(x^k), d^k\rangle$$

$x^{k+1} = x^k + \alpha^k d^k \qquad x^{k+1} - x^k = \alpha^k d^k$

$$\langle \nabla f(x^{k+1}) - \nabla f(x^k), d^k\rangle \leq \|\nabla f(x^{k+1}) - \nabla f(x^k)\| \, \|d^k\|$$

$$\leq L \|\alpha^k d^k\| \, \|d^k\| = \alpha^k L \|d^k\|^2$$

$$\Rightarrow \alpha^k \geq \frac{c_2 - 1}{L} \frac{\langle \nabla f(x^k), d^k\rangle}{\|d^k\|^2}$$

$$\cos\theta^k = \frac{-\langle \nabla f(x^k), d^k\rangle}{\|\nabla f(x^k)\| \, \|d^k\|}$$

$$f(x^k + \alpha d^k) \leq f(x^k) + c_1 \alpha^k \langle \nabla f(x^k), d^k\rangle$$

$$\Rightarrow f(x^{k+1}) \leq f(x^k) + c_1 \frac{c_2 - 1}{L} \frac{\langle \nabla f(x^k), d^k\rangle^2}{\|d^k\|^2}$$

$$\Rightarrow f(x^{k+1}) \leq f(x^k) + c_1 \frac{c_2 - 1}{L} \cos^2\theta^k \|\nabla f(x^k)\|^2$$

再关于k项求和

$$f(x^{k+1}) \leq f(x^0) - c_1 \frac{1 - c_2}{L} \sum_{j=0}^{k} \cos^2\theta^j \|\nabla f(x^j)\|^2$$

$$\Rightarrow c_1 \frac{1 - c_2}{L} \sum_{j=0}^{k} \cos^2\theta^j \|\nabla f(x^j)\|^2 \leq f(x^0) - f(x^{k+1})$$

因为 $f(x^{k+1})$ 有下界，$f(x^{k+1}) > -\infty$

$$\Rightarrow \sum_{j=0}^{k} \cos^2\theta^j \|\nabla f(x^j)\|^2 < +\infty$$

$k \to \infty \Rightarrow \nabla f(x^k) \to 0$

$$k \to +\infty \Rightarrow \sum_{j=0}^{\infty} \cos^2\theta^j \|\nabla f(x^j)\|^2 < +\infty$$

Corollary（线搜索算法收敛性）设 θ^k 是负梯度 $-\nabla f(x^k)$ 和下降方向 d^k 的夹角，并对任意的 k，存在常数 $\gamma>0$，使得

下降方向 d^k 与负梯度方向不能正交

$$\theta^k < \frac{\pi}{2} - \gamma$$

若 Zoutendijk 定理成立，则

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

Proof: 假设结论不成立，即存在子列 $\{k_l\}$ 和正常数 $\delta>0$，使得

$$\|\nabla f(x^{k_l})\| \geq \delta \qquad \forall\, l=1,2,\cdots$$

对任意的 k，有 $\theta^k < \frac{\pi}{2} - \gamma \quad \Rightarrow \quad \theta^k + \gamma < \frac{\pi}{2}$

$$\cos\theta^k > \sin\gamma > 0$$

由于 Zoutendijk 定理成立，有

$$\sum_{j=0}^{\infty} \cos^2\theta^j \|\nabla f(x^j)\|^2 < +\infty$$

$$\sum_{k=0}^{\infty} \cos^2\theta^k \|\nabla f(x^k)\|^2 \geq \sum_{l=0}^{\infty} \cos^2\theta^{k_l} \|\nabla f(x^{k_l})\|^2$$

$$\geq \sum_{l=0}^{\infty} \sin^2\gamma \cdot \delta^2 \to +\infty$$

矛盾，因此我们有

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

算法：带线搜索 梯度下降法

1. 给定 x^0, $\alpha^0 > 0$, 参数 $c, \gamma, \varepsilon \in (0,1)$
2. while $\|\nabla f(x^k)\| > \varepsilon$ do
3. set $\alpha = \alpha^0$
4. while $f(x^k - \alpha \nabla f(x^k)) \geq f(x^k) - c_1 \alpha \|\nabla f(x^k)\|^2$ do
5. 令 $\alpha = \gamma \cdot \alpha$
6. end while
7. 令 $x^{k+1} = x^k - \alpha \nabla f(x^k)$
8. end while
9. output x^{k+1}

for循环
最大迭代步 $M=1000$

for循环
最大迭代步 $m=5$