第三章 非线性规划

第三章 非线性规划

第四节 无约束优化问题的解法

- 最速下降法
- *Newton*法
- 拟*Newton*法
- 共轭梯度法

一. 最速下降法

$$(NP)\ \min_{X\in R^n} f(X)$$

- 收敛性问题的基本概念
- 最速下降法的迭代原理
- 最速下降法的迭代步骤
- 最速下降法的举例
- 最速下降法的收敛结论

1. 收敛性问题的基本概念

定义3-9

若序列$\{X^{(k)}\}$，对于$\forall \varepsilon > 0$，存在正整数$N(\varepsilon)$,

当 $k > N$ 时，有 $\left\| X^{(k)} - X^{*} \right\| < \varepsilon$，即$\left\| X^{(k)} - X^{*} \right\| \xrightarrow[k \to \infty]{} 0$,

则称$\{X^{(k)}\}$收敛于 X^{*}，记为$X^{(k)} \xrightarrow[k \to \infty]{} X^{*}$.

1. 收敛性问题的基本概念

若$\left\| X^{(k)} - X^{*} \right\| \xrightarrow[k\to\infty]{} 0,$

则$X^{(k)} \xrightarrow[k\to\infty]{} X^{*}.$

定义3-10

若 $\{X^{(k)}\}$收敛于 X^{*}，且满足 $\lim\limits_{k\to\infty} \dfrac{\left\| X^{(k+1)} - X^{*} \right\|}{\left\| X^{(k)} - X^{*} \right\|^{p}} = \alpha < \infty,$

则 p 称为 $\{X^{(k)}\}$收敛于 X^{*}的阶。

当 $p = 1$ 时，称为一阶收敛；

当 $p = 2$ 时，称为二阶收敛；

当 $1 < p < 2$ 时，称为超线性收敛；

1. 收敛性问题的基本概念

若$\left\|X^{(k)}-X^*\right\| \xrightarrow[k\to\infty]{} 0$，

则$X^{(k)} \xrightarrow[k\to\infty]{} X^*$.

定义3-10

若$\{X^{(k)}\}$收敛于X^*，且满足$\lim\limits_{k\to\infty}\dfrac{\left\|X^{(k+1)}-X^*\right\|}{\left\|X^{(k)}-X^*\right\|^p}=\alpha<\infty$，

则p称为$\{X^{(k)}\}$收敛于X^*的阶。

当$p=2$时，$\left\|X^{(k+1)}-X^*\right\|$与$\left\|X^{(k)}-X^*\right\|^2$同阶无穷小

当$\alpha=1, k\to\infty$时，$\left\|X^{(k+1)}-X^*\right\| \approx \left\|X^{(k)}-X^*\right\|^2$

$\left\|X^{(k)}-X^*\right\|$	$\left\|X^{(k+1)}-X^*\right\|$	$\left\|X^{(k+2)}-X^*\right\|$	$\left\|X^{(k+3)}-X^*\right\|$	$\left\|X^{(k+4)}-X^*\right\|$
0.1	0.01	0.0001	10^{-8}	10^{-16}

1. 收敛性问题的基本概念

若 $\left\| X^{(k)} - X^{*} \right\| \xrightarrow[k\to\infty]{} 0,$

则 $X^{(k)} \xrightarrow[k\to\infty]{} X^{*}.$

定义3-10

若 $\{X^{(k)}\}$收敛于 X^{*}，且满足 $\lim\limits_{k\to\infty} \dfrac{\left\| X^{(k+1)} - X^{*} \right\|}{\left\| X^{(k)} - X^{*} \right\|^{p}} = \alpha < \infty,$

则p称为 $\{X^{(k)}\}$收敛于 X^{*} 的阶。

当 $p = 1$ 时, $\left\| X^{(k+1)} - X^{*} \right\|$与$\left\| X^{(k)} - X^{*} \right\|$ 同阶无穷小

当 $\alpha = 1, k \to \infty$ 时， $\left\| X^{(k+1)} - X^{*} \right\| \approx \left\| X^{(k)} - X^{*} \right\|$

$\left\| X^{(k)} - X^{*} \right\|$	$\left\| X^{(k+1)} - X^{*} \right\|$	$\left\| X^{(k+2)} - X^{*} \right\|$	$\left\| X^{(k+3)} - X^{*} \right\|$	$\left\| X^{(k+4)} - X^{*} \right\|$
0.1	0.09	0.05	0.02	0.01

1. 收敛性问题的基本概念

若$\left\| X^{(k)} - X^{*} \right\| \xrightarrow[k\to\infty]{} 0$,

则$X^{(k)} \xrightarrow[k\to\infty]{} X^{*}$.

定义3-10

若 $\{X^{(k)}\}$收敛于 X^{*}，且满足 $\lim\limits_{k\to\infty} \dfrac{\left\| X^{(k+1)} - X^{*} \right\|}{\left\| X^{(k)} - X^{*} \right\|^{p}} = \alpha < \infty$,

则p称为 $\{X^{(k)}\}$收敛于 X^{*} 的阶。

当 $p = 1$ 时，称为一阶收敛；

当 $p = 2$ 时，称为二阶收敛；

当 $1 < p < 2$ 时，称为超线性收敛；

1. 收敛性问题的基本概念

定义3-12

若某算法对于任意正定二次目标函数,从任意初始点出发,都能经过有限次迭代达到其极小点,则该算法称为具有**二次终止性**的算法或二次收敛算法.

$$f(x_1, x_2, \cdots, x_n) = \frac{1}{2} X^T Q X + b^T X + c$$

当 Q 为正定阵时，称 $f(X)$ 为正定二次函数。

结论： 正定二次函数 $f(X) = \frac{1}{2} X^T Q X + b^T X + c$ 有唯一全局极小点：$X^* = -Q^{-1} b$

一. 最速下降法

$$(NP)\ \min_{X\in R^n} f(X)$$

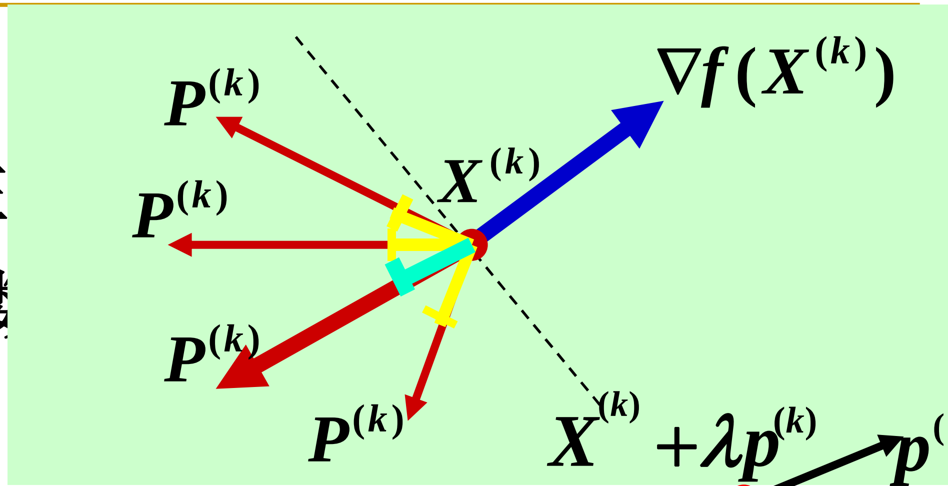
- 收敛性问题的基本概念
- 最速下降法的迭代原理
- 最速下降法的迭代步骤
- 最速下降法的举例
- 最速下降法的收敛结论

2. 迭代原理

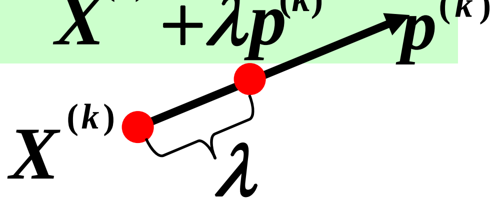
梯度的性质： 函数$f(X)$在

是$X^{(k)}$处函数

证明：



一元函数泰勒公式：



$$f(x^{(k)}+h)=f(x^{(k)})+f'(x^{(k)})h+o(h)$$

$$f(X^{(k)}+\lambda p^{(k)})=f(X^{(k)})+\lambda\nabla f(X^{(k)})^T p^{(k)}+o(\lambda)$$

$$f(X^{(k)}+\lambda p^{(k)})-f(X^{(k)})=\lambda\nabla f(X^{(k)})^T p^{(k)}+o(\lambda)$$

$$=\lambda(\underbrace{\nabla f(X^{(k)})^T p^{(k)}}_{<0}+\frac{o(\lambda)}{\lambda})<0$$

λ充分小时

结论： 当 $\nabla f(X^{(k)})^T p^{(k)}<0$时，$p^{(k)}$是$f(X)$在 $X^{(k)}$处的下降方向。

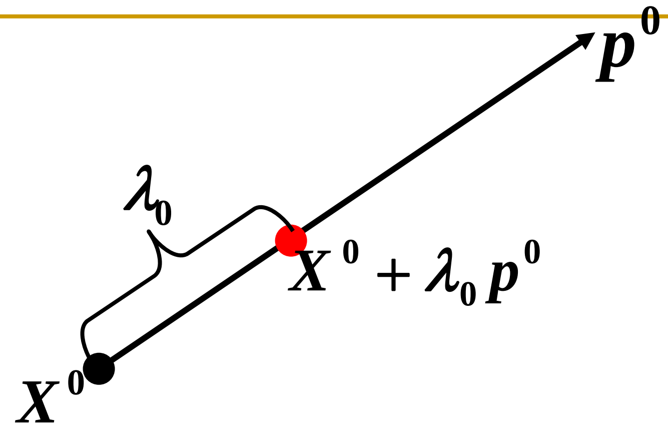
$$\nabla f(X^{(k)})^T p^{(k)}=\left\|\nabla f(X^{(k)})\right\|\left\|p^{(k)}\right\|\cos\theta$$

$$\cos\theta=-1\rightarrow\theta=\pi$$

2. 迭代原理

$$\min_{X\in R^n} f(X)$$

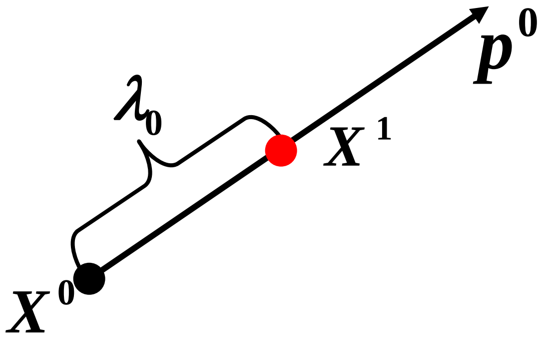
λ_0——最优步长



$$X^0,\quad p^0=-\nabla f(X^0),\ \min_{\lambda\ge 0} f(X^0+\lambda p^0)= f(X^0+\lambda_0 p^0),\ X^1=X^0+\lambda_0 p^0$$

$$f(X^0)>f(X^1)$$

最速下降法迭代原理：



$$\min_{X\in R^n} f(X)= x_1^4 + x_2^2 + 2$$

$$\nabla f(X) = (4x_1^3, 2x_2)^T$$

$$X^0,\quad p^0 = -\nabla f(X^0),\ \min_{\lambda\geq 0} f(X^0+\lambda p^0) = f(X^0+\lambda_0 p^0),\ X^1 = X^0+\lambda_0 p^0$$

$$X^0 = (1,1)^T,\ p^0 = -\nabla f(X^0) = -(4,2)^T$$

$$X^0+\lambda p^0 = \begin{pmatrix}1\\1\end{pmatrix} - \lambda\begin{pmatrix}4\\2\end{pmatrix} = \begin{pmatrix}1-4\lambda\\1-2\lambda\end{pmatrix}$$

$$f(X^0+\lambda p^0) = (1-4\lambda)^4 + (1-2\lambda)^2 + 2 = F(\lambda)$$

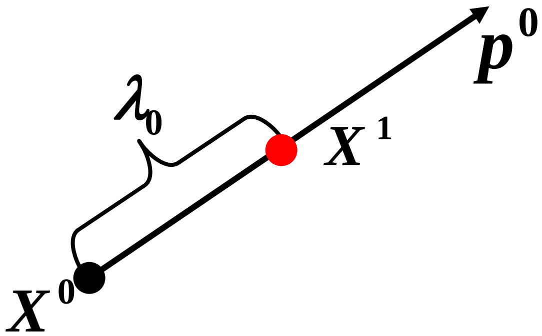
$$\min_{\lambda\geq 0} f(X^0+\lambda p^0) = F(\lambda)$$

λ	$F(\lambda)$
0	4
0.5	3
1	84

一维搜索找极小点 λ_0： 1)确定[0 , 1], 精度0.1

2)用0.618法得到 $\lambda_0 = 0.34375$

最速下降法迭代原理：



$$\min_{X\in R^n} f(X) = x_1^4 + x_2^2 + 2$$

$$\nabla f(X) = (4x_1^3, 2x_2)^T$$

$$X^0,\quad p^0 = -\nabla f(X^0),\ \min_{\lambda\ge 0} f(X^0+\lambda p^0) = f(X^0+\lambda_0 p^0),\ X^1 = X^0+\lambda_0 p^0$$

$$X^0 = (1,1)^T,\ p^0 = -\nabla f(X^0) = -(4,2)^T$$

$$X^0+\lambda p^0 = \begin{pmatrix}1\\1\end{pmatrix} - \lambda\begin{pmatrix}4\\2\end{pmatrix} = \begin{pmatrix}1-4\lambda\\1-2\lambda\end{pmatrix}$$

$$f(X^0+\lambda p^0) = (1-4\lambda)^4 + (1-2\lambda)^2 + 2 = F(\lambda)$$

$$\min_{\lambda\ge 0} f(X^0+\lambda p^0) = F(\lambda) \qquad \lambda_0 = 0.34375$$

$$X^1 = (1,1)^T - 0.34375(4,2)^T = (-0.375, 0.3125)^T$$

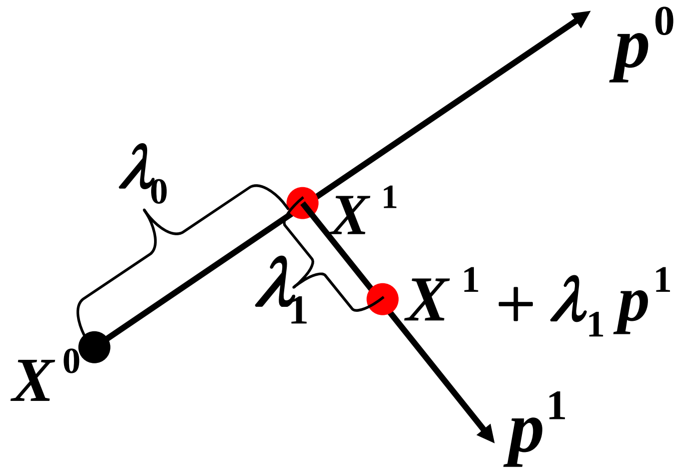
$$f(X_1) = 2.11743 < f(X_0) = 4$$

2. 迭代原理

$$\min_{X \in R^n} f(X)$$

λ_0——最优步长

λ_1——最优步长



$$X^0,\quad p^0=-\nabla f(X^0),\ \min_{\lambda\geq 0} f(X^0+\lambda p^0)=f(X^0+\lambda_0 p^0),\ X^1=X^0+\lambda_0 p^0$$

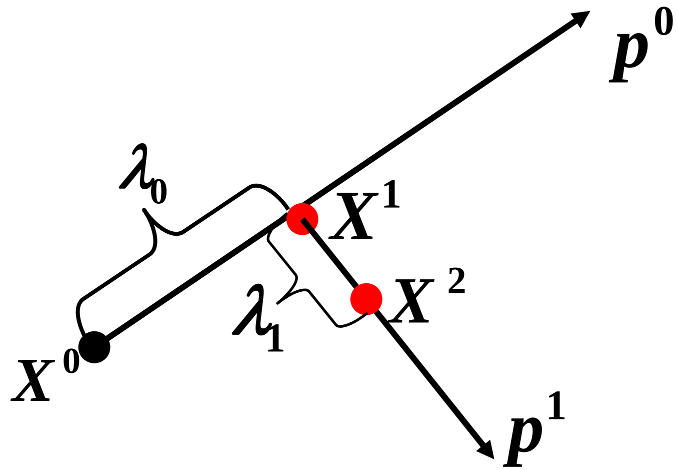
$$X^1,\quad p^1=-\nabla f(X^1),\ \min_{\lambda\geq 0} f(X^1+\lambda p^1)=f(X^1+\lambda_1 p^1),\ X^2=X^1+\lambda_1 p^1$$

$$f(X^0)>f(X^1)>f(X^2)$$

2. 迭代原理

$$\min_{X \in R^n} f(X) \qquad \lambda_0\text{——最优步长}$$

$$\lambda_1\text{——最优步长}$$



$$X^0,\quad p^0=-\nabla f(X^0),\ \min_{\lambda\ge 0} f(X^0+\lambda p^0)=f(X^0+\lambda_0 p^0),\ X^1=X^0+\lambda_0 p^0$$

$$X^1,\quad p^1=-\nabla f(X^1),\ \min_{\lambda\ge 0} f(X^1+\lambda p^1)=f(X^1+\lambda_1 p^1),\ X^2=X^1+\lambda_1 p^1$$

$$X^k,\quad p^k=-\nabla f(X^k),\ \min_{\lambda\ge 0} f(X^k+\lambda p^k)=f(X^k+\lambda_k p^k), X^{k+1}=X^k+\lambda_k p^k$$

得到一个点列：$X^0, X^1, \cdots, X^{(k)}, \cdots,$

可以证明：$f(X^{(k)})$严格 $\downarrow$, $\therefore X^{(k)} \xrightarrow[k\to\infty]{} X^* \ (Th3-10),$

线性收敛 $\qquad (Th3-11)$

$$f(X^0)>f(X^1)>f(X^2)>\cdots>f(X^k)>f(X^{k+1})$$

2. 迭代原理 $\min\limits_{X\in R^n} f(X)$

$$X^0,\quad p^0=-\nabla f(X^0),\ \min_{\lambda\ge 0} f(X^0+\lambda p^0)= f(X^0+\lambda_0 p^0),\ X^1=X^0+\lambda_0 p^0$$

$$X^1,\quad p^1=-\nabla f(X^1),\ \min_{\lambda\ge 0} f(X^1+\lambda p^1)= f(X^1+\lambda_1 p^1),\ X^2=X^1+\lambda_1 p^1$$

$$X^k,\quad p^k=-\nabla f(X^k),\ \min_{\lambda\ge 0} f(X^k+\lambda p^k)= f(X^k+\lambda_k p^k),X^{k+1}=X^k+\lambda_k p^k$$

得到一个点列 $:X^0,X^1,\cdots,X^{(k)},\cdots,$

证明： $f(X^{(k)})$严格$\downarrow$

$$\because \nabla f(X^{(k)})^T p^{(k)}=-\nabla f(X^{(k)})^T\nabla f(X^{(k)})=-\left\|\nabla f(X^{(k)})\right\|^2<0$$

$$f(X^{(k)}+\lambda p^{(k)})=f(X^{(k)})+\lambda\nabla f(X^{(k)})^T p^{(k)}+o(\lambda)$$

$$f(X^{(k)}+\lambda p^{(k)})-f(X^{(k)})=\lambda\nabla f(X^{(k)})^T p^{(k)}+o(\lambda)$$

$$=\lambda(\underbrace{\nabla f(X^{(k)})^T p^{(k)}}_{<0}+\frac{o(\lambda)}{\lambda})<0$$

$\lambda>0$充分小时

$$\therefore f(\overbrace{X^{(k)}+\lambda_k p^{(k)}}^{X^{k+1}})-f(X^{(k)})<0\ \therefore\ f(X^{(k)})\text{严格}\downarrow,\therefore\ X^{(k)}\xrightarrow[k\to\infty]{}X^*$$

一. 最速下降法

$$(NP)\ \min_{X \in R^n} f(X)$$

- 收敛性问题的基本概念
- 最速下降法的迭代原理
- 最速下降法的迭代步骤
- 最速下降法的举例
- 最速下降法的收敛结论

3. 迭代步骤

1^0 取初始点 $X^{(0)}$,容许误差(精度)$\varepsilon > 0$,令$k := 0$

2^0 计算 $p^{(k)} = -\nabla f(X^{(k)})$

3^0 检验$\left\| p^{(k)} \right\| \le \varepsilon$?若是迭代终止,取$X^* = X^{(k)}$,否则转 4^0

4^0 求最优步长 λ_k :$\min\limits_{\lambda \ge 0} f(X^{(k)} + \lambda p^{(k)}) = f(X^{(k)} + \lambda_k p^{(k)})$(一维搜索)

5^0 令$X^{(k+1)} = X^{(k)} + \lambda_k p^{(k)}$,令$k := k+1$,转$2^0$

$k = 0 \qquad p^0 = -\nabla f(X^0)$, $\left\| p^{(0)} \right\| \le \varepsilon$?若是取 $X^* = X^{(0)}$,否则

$\min\limits_{\lambda \ge 0} f(X^0 + \lambda p^0) = f(X^0 + \lambda_0 p^0)$, $\quad X^1 = X^0 + \lambda_0 p^0$

$k = 1 \quad p^1 = -\nabla f(X^1)$, $\left\| p^{(1)} \right\| \le \varepsilon$?若是取 $X^* = X^{(1)}$,否则

$\min\limits_{\lambda \ge 0} f(X^1 + \lambda p^1) = f(X^1 + \lambda_1 p^1)$, $\quad X^2 = X^1 + \lambda_1 p^1$

$k = 2 \quad p^2 = -\nabla f(X^2)$, $\left\| p^{(2)} \right\| \le \varepsilon$?若是取 $X^* = X^{(2)}$,否则

$\min\limits_{\lambda \ge 0} f(X^2 + \lambda p^2) = f(X^2 + \lambda_2 p^2)$, $\quad X^3 = X^2 + \lambda_2 p^2$

3. 迭代步骤

$$\nabla f(X^{(k)}) = (\frac{\partial f}{\partial x_1}(X^{(k)}), \frac{\partial f}{\partial x_2}(X^{(k)}), \cdots, \frac{\partial f}{\partial x_n}(X^{(k)}))^T$$

$$\downarrow$$

$$\nabla f(X^{*}) = (\frac{\partial f}{\partial x_1}(X^{*}), \frac{\partial f}{\partial x_2}(X^{*}), \cdots, \frac{\partial f}{\partial x_n}(X^{*}))^T$$

1^0 取初始点 $X^{(0)}$，…

2^0 计算 $p^{(k)} = -\nabla f($…

3^0 检验 $\|p^{(k)}\| \le \varepsilon$? 若是迭代终止，取 $X^{*} = X^{(k)}$，否则转 4^0

4^0 求最优步长 λ_k : $\min\limits_{\lambda \ge 0} f(X^{(k)} + \lambda p^{(k)}) = f(X^{(k)} + \lambda_k p^{(k)})$（一维搜索）

5^0 令 $X^{(k+1)} = X^{(k)} + \lambda_k p^{(k)}$，令 $k := k+1$，转 2^0

注释：

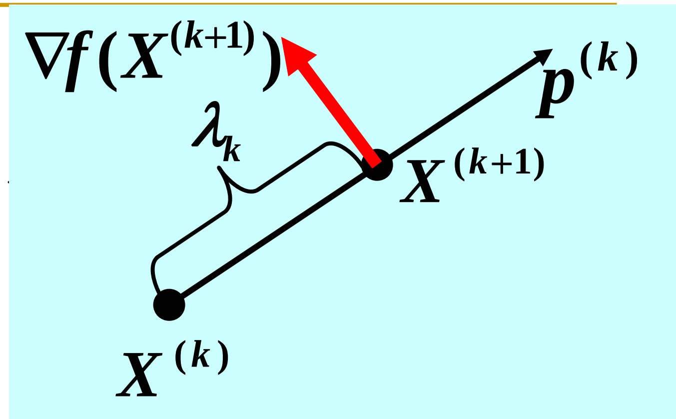
1^0 停机准则： 设 $\nabla f(X)$ 连续（即 $f(X)$ 连续可微）　$\min\limits_{X \in R^n} f(X)$

当 $X^{(k)} \to X^{*}$ 时，　$\nabla f(X^{(k)}) \to \nabla f(X^{*}) = 0$ （一阶必要条件）

$$\|\nabla f(X^{(k)})\| \to \|\nabla f(X^{*})\| = 0$$

$$\|\nabla f(X^{(k)})\| < \varepsilon \longrightarrow \|p^{(k)}\| < \varepsilon$$

3. 迭代步骤



1^0取初始点 $X^{(0)}$,容许误差(精度)$\varepsilon > 0,$

2^0计算$p^{(k)} = -\nabla f(X^{(k)})$

3^0检验$\|p^{(k)}\| \le \varepsilon$?若是迭代终止，取$X^* =$

4^0求最优步长λ_k,$\min\limits_{\lambda \ge 0} f(X^{(k)} + \lambda p^{(k)}) = f(X^{(k)} + \lambda_k p^{(k)})$(一维搜索)

5^0令$X^{(k+1)} = X^{(k)} + \lambda_k p^{(k)}$,令$k := k+1$,转$2^0$

注释：

2^0 结论：一维搜索最优解的梯度 $\nabla f(X^{(k+1)})$ 与搜索方向 $p^{(k)}$正交

证明：

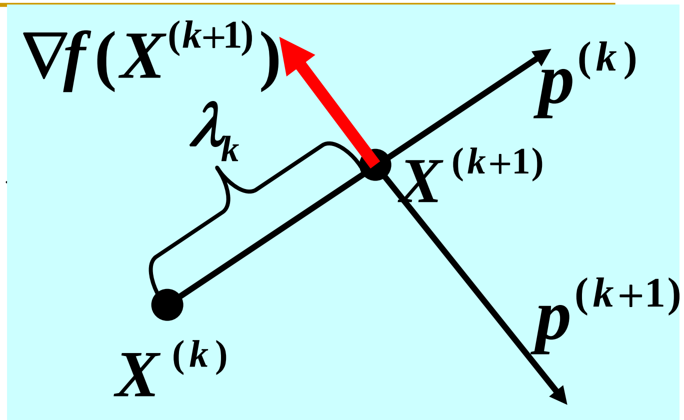
$$\because \min_{\lambda \ge 0} \underbrace{f(X^{(k)} + \lambda p^{(k)})}_{\varphi(\lambda)} = \underbrace{f(X^{(k)} + \lambda_k p^{(k)})}_{\varphi(\lambda_k)}$$

$$\varphi'(\lambda) = \nabla f(X^{(k)} + \lambda p^{(k)})^T p^{(k)}$$

$$\varphi'(\lambda_k) = \nabla f(X^{(k)} + \lambda_k p^{(k)})^T p^{(k)}$$

$$\therefore 0 = \varphi'(\lambda_k) = \nabla f(X^{(k+1)})^T p^{(k)} \quad (3-11)$$

3. 迭代步骤



1^0 取初始点 $X^{(0)}$, 容许误差 (精度) $\varepsilon > 0$,

2^0 计算 $p^{(k)} = -\nabla f(X^{(k)})$

3^0 检验 $\|p^{(k)}\| \le \varepsilon$? 若是迭代终止，取 $X^* =$

4^0 求最优步长 λ_k, $\min\limits_{\lambda \ge 0} f(X^{(k)} + \lambda p^{(k)}) = f(X^{(k)} + \lambda_k p^{(k)})$(一维搜索)

5^0 令 $X^{(k+1)} = X^{(k)} + \lambda_k p^{(k)}$, 令 $k := k+1$, 转 2^0

注释:

3^0 结论: 最速下降法的任何两个相邻搜索方向正交(垂直)

$$\because \min_{\lambda\ge 0} \underbrace{f(X^{(k)}+\lambda p^{(k)})}_{\varphi(\lambda)} = \underbrace{f(X^{(k)}+\lambda_k p^{(k)})}_{\varphi(\lambda_k)} \quad \varphi'(\lambda) = \nabla f(X^{(k)}+\lambda p^{(k)})^T p^{(k)}$$

$$\boxed{\therefore 0 = \varphi'(\lambda_k) = \nabla f(X^{(k+1)})^T p^{(k)}} \ (3-11) \longrightarrow p^{(k+1)^T} p^{(k)} = 0$$

3. 迭代步骤

1^0 取初始点 $X^{(0)}$，容许误差（精度）$\varepsilon > 0$，令 $k := 0$

2^0 计算 $p^{(k)} = -\nabla f(X^{(k)})$

3^0 检验 $\left\| p^{(k)} \right\| \le \varepsilon$? 若是迭代终止，取 $X^* = X^{(k)}$，否则转 4^0

4^0 求最优步长 λ_k，$\min\limits_{\lambda \ge 0} f(X^{(k)} + \lambda p^{(k)}) = f(X^{(k)} + \lambda_k p^{(k)})$（一维搜索）

5^0 令 $X^{(k+1)} = X^{(k)} + \lambda_k p^{(k)}$，令 $k := k+1$，转 2^0

注释：

4^0 将一维搜索用于正定二次函数：$f(X) = \frac{1}{2} X^T Q X + b^T X + c$

则可以得到 λ_k 的表达式：　　令 $g^{(k)} = \nabla f(X^{(k)})$

$$\boxed{\lambda_k = -\frac{{g^{(k)}}^T p^{(k)}}{{p^{(k)}}^T Q p^{(k)}}} \quad (3-13)$$

3. 迭代步骤

4⁰ 将一维搜索用于正定二次函数：$f(X)=\frac{1}{2}X^TQX+b^TX+c$

则可以得到 λ_k 的表达式：　令 $g^{(k)}=\nabla f(X^{(k)})$

$$\boxed{\lambda_k=-\frac{g^{(k)T}p^{(k)}}{p^{(k)T}Qp^{(k)}}} \quad (3-13)$$

$$\nabla f(X)=QX+b$$

$$X^{(k+1)}=X^{(k)}+\lambda_k p^{(k)}$$

$$\nabla f(X^{(k+1)})^T p^{(k)}=0$$

证明：

$$g^{(k)}=\nabla f(X^{(k)})=QX^{(k)}+b$$

$$g^{(k+1)}=\nabla f(X^{(k+1)})=QX^{(k+1)}+b=Q(X^{(k)}+\lambda_k p^{(k)})+b$$

$$=QX^{(k)}+\lambda_k Qp^{(k)}+b=g^{(k)}+\lambda_k Qp^{(k)}$$

$$\because 0=g^{(k+1)T}p^{(k)}=(g^{(k)}+\lambda_k Qp^{(k)})^T p^{(k)}=g^{(k)T}p^{(k)}+\lambda_k p^{(k)T}Qp^{(k)}$$

$$\therefore \lambda_k=-\frac{g^{(k)T}p^{(k)}}{p^{(k)T}Qp^{(k)}}$$

■ **注释：**该公式具有普遍性

3. 迭代步骤

$$\min_{X\in R^n} f(X)$$

1^0 取初始点 $X^{(0)}$,容许误差(精度)$\varepsilon>0$,令$k:=0$

2^0 计算$p^{(k)}=-\nabla f(X^{(k)})$

3^0 检验$\left\|p^{(k)}\right\|\le\varepsilon$?若是迭代终止,取$X^*=X^{(k)}$,否则转 4^0

4^0 求最优步长 λ_k,$\min\limits_{\lambda\ge0} f(X^{(k)}+\lambda p^{(k)})=f(X^{(k)}+\lambda_k p^{(k)})$(一维搜索)

5^0 令$X^{(k+1)}=X^{(k)}+\lambda_k p^{(k)}$,令$k:=k+1$,转$2^0$

注释:

4

4^0 将一维搜索用于正定二次函数:$f(X)=\frac{1}{2}X^TQX+b^TX+c$

则可以得到 λ_k 的表达式: 令$g^{(k)}=\nabla f(X^{(k)})$

$$\lambda_k=-\frac{g^{(k)^T}p^{(k)}}{p^{(k)^T}Qp^{(k)}}\quad(3-13)$$

3. 迭代步骤

1^0 取初始点 $X^{(0)}$，容许误差 (精度) $\varepsilon > 0$，令 $k := 0$

2^0 计算 $p^{(k)} = -\nabla f(X^{(k)})$

3^0 检验 $\left\| p^{(k)} \right\| \leq \varepsilon$？若是迭代终止，取 $X^* = X^{(k)}$，否则转 4^0

4^0 求最优步长 λ_k，$\min\limits_{\lambda \geq 0} f(X^{(k)} + \lambda p^{(k)}) = f(X^{(k)} + \lambda_k p^{(k)})$（一维搜索）

5^0 令 $X^{(k+1)} = X^{(k)} + \lambda_k p^{(k)}$，令 $k := k+1$，转 2^0

注释：

5^0 将最速下降法用于正定二次函数：$f(X) = \frac{1}{2} X^T Q X + b^T X + c$

则可以得到 λ_k 的表达式：　　令 $g^{(k)} = \nabla f(X^{(k)})$

$$\lambda_k = -\frac{g^{(k)^T} p^{(k)}}{p^{(k)^T} Q p^{(k)}} \quad (3-13) \xrightarrow{p^{(k)} = -g^{(k)}} \lambda_k = \frac{g^{(k)^T} g^{(k)}}{g^{(k)^T} Q g^{(k)}} \quad (3-14)$$

3. 迭代步骤

1^0 取初始点 $X^{(0)}$，容许误差（精度）$\varepsilon > 0$，令 $k := 0$

2^0 计算 $p^{(k)} = -\nabla f(X^{(k)})$

3^0 检验 $\left\| p^{(k)} \right\| \le \varepsilon$？若是迭代终止，取 $X^* = X^{(k)}$，否则转 4^0

4^0 求最优步长 λ_k，$\min\limits_{\lambda \ge 0} f(X^{(k)} + \lambda p^{(k)}) = f(X^{(k)} + \lambda_k p^{(k)})$（一维搜索）

5^0 令 $X^{(k+1)} = X^{(k)} + \lambda_k p^{(k)}$，令 $k := k+1$，转 2^0

注释：

5^0 最速下降法，*Newton*法，拟*Newton*法，共轭梯度法的区别

就是搜索方向 $p^{(k)}$ 取得不同。

$$p^{(k)} = -\nabla f(X^{(k)})$$

$$p^{(k)} = -G(X^{(k)})^{-1} \nabla f(X^{(k)})$$

$$p^{(k)} = -H^{(k)} \nabla f(X^{(k)})$$

$$p^{(k)} = -\nabla f(X^{(k)}) + \beta_{k-1} p^{(k-1)}$$

一. 最速下降法

$$(NP)\quad \min_{X \in R^n} f(X)$$

- 收敛性问题的基本概念
- 最速下降法的迭代原理
- 最速下降法的迭代步骤
- 最速下降法的举例
- 最速下降法的收敛结论

$$f(X)=\frac{1}{2}X^TQX+b^TX+c$$

4. 举例

例3-10 用最速下降法求 $f(X)=x_1^2+4x_2^2$ 的极小点，迭代两次。$X^{(0)}=(1,1)^T, \varepsilon=10^{-4}$ $X^*=(0,0)^T$

解:

4. 举例

$$f(X)=\frac{1}{2}X^TQX+b^TX+c$$

例3-10 用最速下降法求 $f(X)=x_1^2+4x_2^2$ 的极小点，迭代两次。$X^{(0)}=(1,1)^T,\varepsilon=10^{-4}$

解：

$$f(X)=\frac{1}{2}(2x_1^2+8x_2^2)=\frac{1}{2}X^TQX \quad Q=\begin{pmatrix}2 & 0\\ 0 & 8\end{pmatrix} \quad g(X)=\nabla f(X)=\begin{pmatrix}2x_1\\ 8x_2\end{pmatrix}$$

$2^0\ p^{(k)}=-\nabla f(X^{(k)})\quad X^{(0)}=(1,1)^T,\varepsilon=10^{-4}$

$3^0\ \left\|p^{(k)}\right\|\le\varepsilon$?若是,$X^*=X^{(k)}$,否则转$4^0$

4^0求λ_k,$\min\limits_{\lambda\ge 0} f(X^{(k)}+\lambda p^{(k)})=f(X^{(k)}+\lambda_k p^{(k)})$(一维搜索)　$4^0\ \lambda_k=-\dfrac{g^{(k)T}p^{(k)}}{p^{(k)T}Qp^{(k)}}$

$5^0\ X^{(k+1)}=X^{(k)}+\lambda_k p^{(k)},k:=k+1$,转$2^0$

解:

$$f(X)=\frac{1}{2}(2x_1^2+8x_2^2)=\frac{1}{2}X^TQX\quad Q=\begin{pmatrix}2&0\\0&8\end{pmatrix}\quad g(X)=\nabla f(X)=\begin{pmatrix}2x_1\\8x_2\end{pmatrix}$$

1　$k=0$

$(2)p^{(0)}=-g^{(0)}=-\begin{pmatrix}2\\8\end{pmatrix}$　　$(3)\left\|p^{(0)}\right\|=\sqrt{(-2)^2+(-8)^2}=\sqrt{68}$(太大)

$$(4)\lambda_0=-\frac{g^{(0)T}p^{(0)}}{p^{(0)T}Qp^{(0)}}=\frac{(2,8)\begin{pmatrix}2\\8\end{pmatrix}}{(2,8)\begin{pmatrix}2&0\\0&8\end{pmatrix}\begin{pmatrix}2\\8\end{pmatrix}}=\frac{68}{520}=0.13077$$

$2^0\ p^{(k)} = -\nabla f(X^{(k)})$

$3^0\ \left\| p^{(k)} \right\| \le \varepsilon$?若是,$X^* = X^{(k)}$,否则转$4^0$

4^0求λ_k,$\min\limits_{\lambda\ge 0} f(X^{(k)} + \lambda p^{(k)}) = f(X^{(k)} + \lambda_k p^{(k)})$(一维搜索) $4^0\ \lambda_k = -\dfrac{g^{(k)T} p^{(k)}}{p^{(k)T} Q p^{(k)}}$

$5^0\ X^{(k+1)} = X^{(k)} + \lambda_k p^{(k)}, k := k+1$,转$2^0$

解:

$$f(X) = \frac{1}{2}(2x_1^2 + 8x_2^2) = \frac{1}{2} X^T Q X \quad Q = \begin{pmatrix} 2 & 0 \\ 0 & 8 \end{pmatrix} \quad g(X) = \nabla f(X) = \begin{pmatrix} 2x_1 \\ 8x_2 \end{pmatrix}$$

1 $k=0$

$$(2)\, p^{(0)} = -g^{(0)} = -\begin{pmatrix} 2 \\ 8 \end{pmatrix} \qquad (3)\left\| p^{(0)} \right\| = \sqrt{(-2)^2 + (-8)^2} = \sqrt{68}\text{(太大)}$$

$$(4)\,\lambda_0 = 0.13077$$

$$(5)\, X^{(1)} = X^{(0)} + \lambda_0 p^{(0)} = \begin{pmatrix} 1 \\ 1 \end{pmatrix} - 0.13077 \begin{pmatrix} 2 \\ 8 \end{pmatrix} = \begin{pmatrix} 0.73846 \\ -0.04616 \end{pmatrix} \qquad k=1$$

2^0 $p^{(k)} = -\nabla f(X^{(k)})$

3^0 $\left\| p^{(k)} \right\| \le \varepsilon$?若是，$X^* = X^{(k)}$，否则转$4^0$

4^0 求λ_k，$\min_{\lambda \ge 0} f(X^{(k)} + \lambda p^{(k)}) = f(X^{(k)} + \lambda_k p^{(k)})$（一维搜索） 4^0 $\lambda_k = -\dfrac{g^{(k)T} p^{(k)}}{p^{(k)T} Q p^{(k)}}$

5^0 $X^{(k+1)} = X^{(k)} + \lambda_k p^{(k)}$，$k := k+1$，转$2^0$

解：

$$f(X) = \frac{1}{2}(2x_1^2 + 8x_2^2) = \frac{1}{2} X^T Q X \quad Q = \begin{pmatrix} 2 & 0 \\ 0 & 8 \end{pmatrix} \quad g(X) = \nabla f(X) = \begin{pmatrix} 2x_1 \\ 8x_2 \end{pmatrix}$$

2 $k = 1$

$$X^{(1)} = \begin{pmatrix} 0.73846 \\ -0.04616 \end{pmatrix} \quad (2) p^{(1)} = -g^{(1)} = -\begin{pmatrix} 1.47692 \\ -0.39623 \end{pmatrix} \quad (3) \left\| p^{(1)} \right\| = 1.52237$$

$$(4) \lambda_1 = -\frac{g^{(1)T} p^{(1)}}{p^{(1)T} Q p^{(1)}} = 0.425$$

$$(5) X^{(2)} = X^{(1)} + \lambda_1 p^{(1)} = \begin{pmatrix} 0.73846 \\ -0.04616 \end{pmatrix} - 0.425 \begin{pmatrix} 1.47692 \\ -0.39623 \end{pmatrix} = \begin{pmatrix} 0.11076 \\ 0.11076 \end{pmatrix} \quad k = 2$$

2^0 $p^{(k)} = -\nabla f(X^{(k)})$

3^0 $\left\| p^{(k)} \right\| \le \varepsilon$?若是, $X^* = X^{(k)}$,否则转4^0

4^0 求λ_k, $\min\limits_{\lambda \ge 0} f(X^{(k)} + \lambda p^{(k)}) = f(X^{(k)} + \lambda_k p^{(k)})$(一维搜索) 4^0 $\lambda_k = \dfrac{g^{(k)T} g^{(k)}}{g^{(k)T} Q g^{(k)}}$

5^0 $X^{(k+1)} = X^{(k)} + \lambda_k p^{(k)}, k := k+1$,转$2^0$

解:

$$f(X) = \frac{1}{2}(2x_1^2 + 8x_2^2) = \frac{1}{2} X^T Q X \quad Q = \begin{pmatrix} 2 & 0 \\ 0 & 8 \end{pmatrix} \quad g(X) = \nabla f(X) = \begin{pmatrix} 2x_1 \\ 8x_2 \end{pmatrix}$$

3 $k = 2$

$$X^{(2)} = \begin{pmatrix} 0.11076 \\ 0.11076 \end{pmatrix} \qquad (2)\ p^{(2)} = -g^{(2)} = -\begin{pmatrix} 0.22152 \\ 0.88608 \end{pmatrix}$$

(3) $\left\| p^{(2)} \right\| = 0.91335$（太大）继续迭代。

注释：最速下降法收敛速度很慢。

4. 举例

例3-10 用最速下降法求 $f(X)=x_1^2+4x_2^2$ 的极小点，迭代两次。 $X^{(0)}=(1,1)^T, \varepsilon=10^{-4}$

注释：

本例的计算结果如图3-14(*P*156).迭代点在向极小点靠近的过程中形成一条锯齿折线,这种现象称为锯齿现象.这是由于最速下降法的任何两个相邻搜索方向正交.因此,从直观上可以看到,在远离极小点的地方,每次迭代可使目标函数值有较大的下降,但越接近极小点,由于锯齿现象,函数值下降速度显著变慢.

优点： 计算简单,存储量小.

缺点： 由于锯齿现象,迭代后期收敛速度变慢.

一. 最速下降法

$$(NP)\quad \min_{X \in R^n} f(X)$$

- 收敛性问题的基本概念
- 最速下降法的迭代原理
- 最速下降法的迭代步骤
- 最速下降法的举例
- 最速下降法的收敛结论

5. 最速下降法的收敛结论

最速下降法所产生的迭代点列

$$X^{(k)} \xrightarrow[k \to \infty]{} X^{*} \ (\text{定理}3-10)$$

线性收敛 $\quad (\text{定理}3-11)$

X^{*}是 $\min\limits_{X \in R^n} f(X)$ 的局部最优解

一. 最速下降法

$$(NP)\ \min_{X\in R^n} f(X)$$

- 收敛性问题的基本概念
- 最速下降法的迭代原理
- 最速下降法的迭代步骤
- 最速下降法的举例
- 最速下降法的收敛结论

作业：P245　14

作业：P155　14