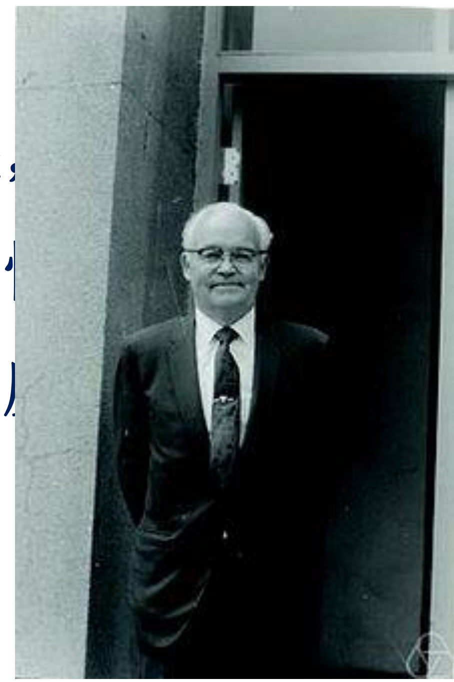
共轭梯度法

一、线性共轭方向法

共轭梯度法： Hestenes、Stiefel (1952)在求解大规模线性方程组时创立的一种迭代算法。



后来，964

非线创立

轭梯

Fletcher

Hestenes

Stiefel

问题引出

线性方程组

对称正定

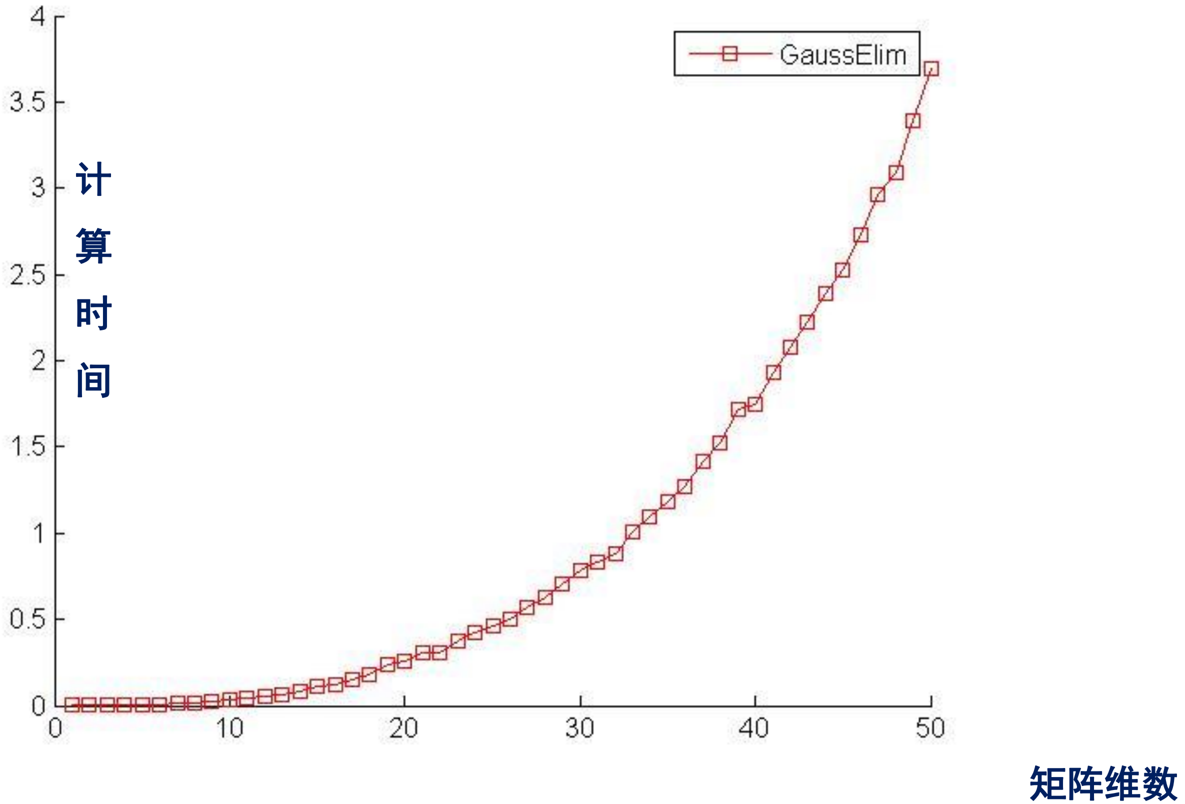
$$\boxed{\boldsymbol{A}}\boldsymbol{x}=\boldsymbol{b}$$

经典算法：

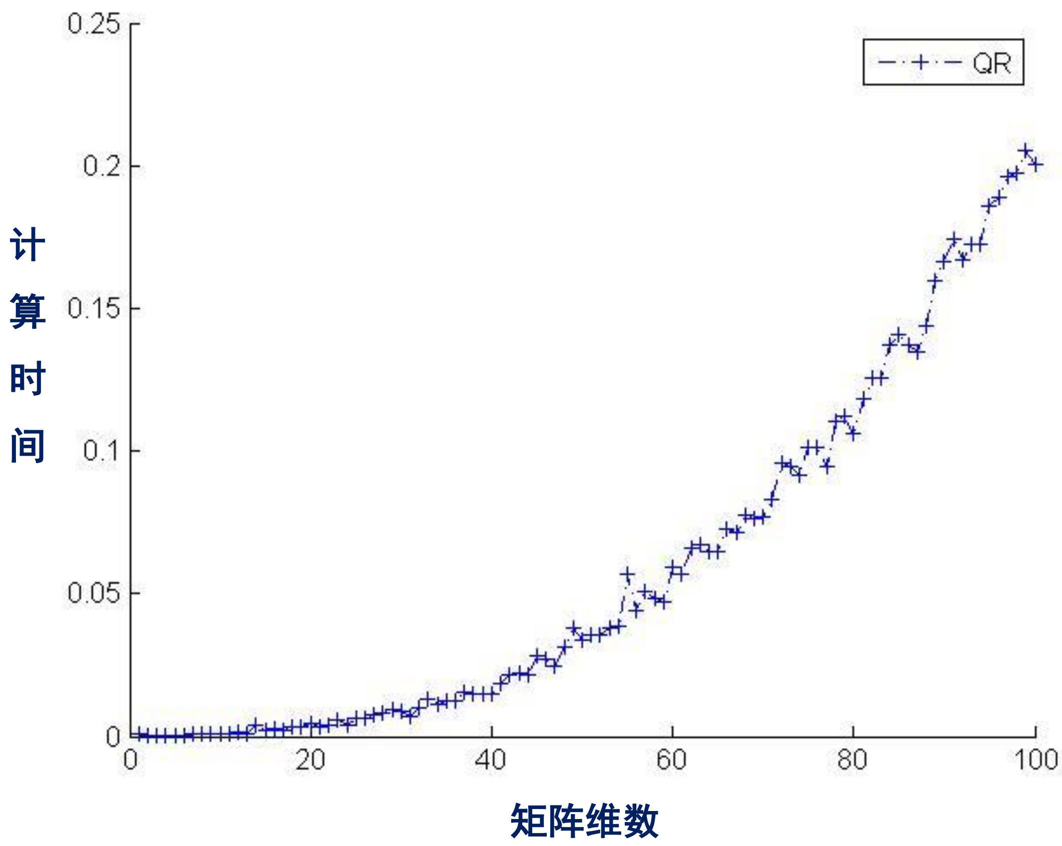
Gauss消元法、系数矩阵三角分解法

算法缺陷： 计算时间随问题规模急速增长。

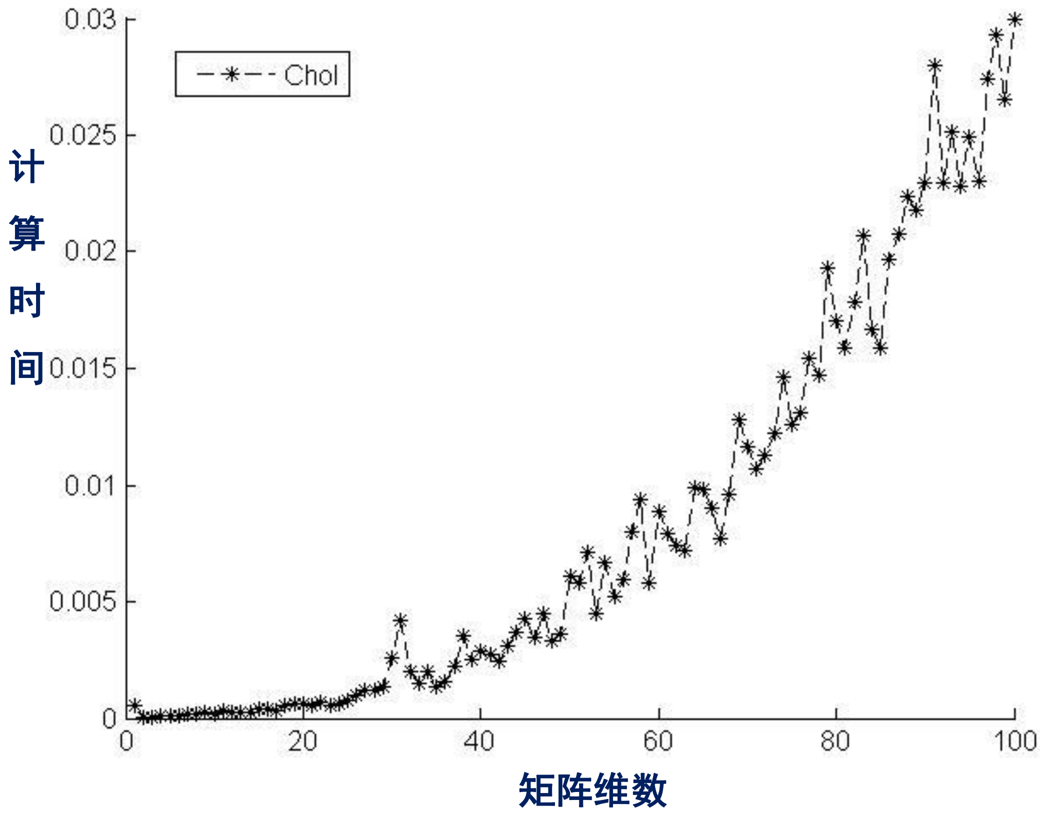
Gauss消元法



系数矩阵QR分解法



系数矩阵Cholesky分解法



问题：如何求解大规模对称线性方程组问题？

$$\underbrace{A}_{\text{对称正定}}x = b$$

利用系数矩阵性质

严格凸二次规划

$$\min_x \frac{1}{2}x^T Ax - b^T x$$

最速下降算法

太慢！

牛顿算法

太累！

能否利用问题结构，构造新算法？

线性共轭方向法

线性共轭方向法

共轭方向： 对对称正定阵 $\boldsymbol{A}$，若 $\boldsymbol{d}_1, \boldsymbol{d}_2 \in \mathbb{R}^n$ 满足 $\boldsymbol{d}_1^{\mathrm{T}} \boldsymbol{A} \boldsymbol{d}_2 = 0$

则称 $\boldsymbol{d}_1$，$\boldsymbol{d}_2$ 关于矩阵 $\boldsymbol{A}$ 共轭，并称其为关于 $\boldsymbol{A}$ 的共轭方向.

共轭是正交的推广.

线性共轭方向的推广

若向量组 $\boldsymbol{d}_1, \boldsymbol{d}_2, \cdots, \boldsymbol{d}_n$ 关于对称正定阵 $\boldsymbol{A}$ 两两共轭，

即满足

$$\boldsymbol{d}_i^{\mathrm{T}} \boldsymbol{A} \boldsymbol{d}_j = 0,\ 1 \leqslant i \neq j \leqslant k,$$

则称该向量组关于矩阵 $\boldsymbol{A}$ 共轭。

性质 若向量组 $\boldsymbol{d}_1, \boldsymbol{d}_2, \cdots, \boldsymbol{d}_n$ 关于矩阵 $\boldsymbol{A}$ 共轭，则它们线性无关.

证明： 利用线性无关的定义。

线性共轭方向法

$$\min_{\boldsymbol{x}\in\mathbb{R}^n} f(\boldsymbol{x}) = \frac{1}{2}\boldsymbol{x}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{x} - \boldsymbol{b}^{\mathrm{T}}\boldsymbol{x}$$

步1、初始点 $\boldsymbol{x}_0$，搜索方向 $\boldsymbol{d}_0$ 满足 $\langle \boldsymbol{d}_0, \boldsymbol{g}_0\rangle < 0$ ，终止参数 $\varepsilon \geqslant 0$ 令 $k = 0$.

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

步3、计算最优步长 $\alpha_k = \arg\min_{\alpha\geqslant 0}\{f(\boldsymbol{x}_k + \alpha \boldsymbol{d}_k)\}$

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

步4、构造 $\boldsymbol{d}_{k+1}$ ，使其与 $\boldsymbol{d}_0, \boldsymbol{d}_1, \cdots, \boldsymbol{d}_k$ 关于矩阵 $\boldsymbol{A}$ 共轭.

令 $k = k + 1$ 返回步2.

二次终止性

定理 对严格凸二次函数 $f(\boldsymbol{x})=\frac{1}{2}\boldsymbol{x}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{x}-\boldsymbol{b}^{\mathrm{T}}\boldsymbol{x}$，向量组 $\boldsymbol{d}_0, \boldsymbol{d}_1, \cdots, \boldsymbol{d}_{n-1}$ 关于 $\boldsymbol{A}$ 共轭。共轭方向法产生点列 $\{\boldsymbol{x}_k\}$。则对任意 $0 \leqslant k \leqslant n-1$, $\boldsymbol{x}_{k+1}$ 是目标函数在仿射集 $\boldsymbol{x}_0+\mathrm{span}[\boldsymbol{d}_0, \boldsymbol{d}_1, \cdots, \boldsymbol{d}_k]$ 上的最小值点, 算法至多 n 步迭代后终止.

有关定义

子空间 $\mathrm{span}[\boldsymbol{d}_0, \boldsymbol{d}_1, \boldsymbol{d}_1, \cdots, \boldsymbol{d}_k]=\{\sum_{i=0}^{k}\alpha_i \boldsymbol{d}_i \mid \alpha_i \in R,\ i=0,1,\cdots,k\}$

仿射集 将子空间平移后得到的集合

$$\boldsymbol{x}_0+\mathrm{span}[\boldsymbol{d}_0, \boldsymbol{d}_1, \boldsymbol{d}_1, \cdots, \boldsymbol{d}_k]=\{\boldsymbol{x}_0+\sum_{i=0}^{k}\alpha_i \boldsymbol{d}_i \mid \alpha_i \in R,\ i=0,1,\cdots,k\}$$

二、线性共轭梯度法

问题： 线性共轭方向法具有二次终止性，如何构造共轭方向？

策略： 取 $d_0 = -g_0$。然后依次以负梯度为基准，对其进行校正，使新搜索方向与已有的搜索方向共轭。

$d_0 = -g_0;$

待定参数

待定参数

待定参数

待定参数

$$\boxed{d_i^{\mathrm{T}} A d_j = 0,\ 1 \leqslant i \neq j \leqslant k,}$$

待定参数

待定参数

共轭方向的构造

严格凸二次函数 $f(\boldsymbol{x}) = \frac{1}{2}\boldsymbol{x}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{x} - \boldsymbol{b}^{\mathrm{T}}\boldsymbol{x}$,

令 $\boldsymbol{d}_0 = -\boldsymbol{g}_0$, $\quad \boldsymbol{x}_1 = \boldsymbol{x}_0 + \alpha_0\boldsymbol{d}_0$,

最优步长

令 $\boldsymbol{d}_1 = -\boldsymbol{g}_1 + \beta_0\boldsymbol{d}_0$,

待定参数

选择参数 β_0满足共轭条件 $\boldsymbol{d}_1^{\mathrm{T}}\boldsymbol{A}\boldsymbol{d}_0 = 0$

解之得 $$\beta_0 = \frac{\boldsymbol{g}_1^{\mathrm{T}}\boldsymbol{A}\boldsymbol{d}_0}{\boldsymbol{d}_0^{\mathrm{T}}\boldsymbol{A}\boldsymbol{d}_0} = \frac{\boldsymbol{g}_1^{\mathrm{T}}(\boldsymbol{g}_1 - \boldsymbol{g}_0)}{\boldsymbol{d}_0^{\mathrm{T}}(\boldsymbol{g}_1 - \boldsymbol{g}_0)} = \frac{\boldsymbol{g}_1^{\mathrm{T}}\boldsymbol{g}_1}{\boldsymbol{g}_0^{\mathrm{T}}\boldsymbol{g}_0}.$$

由此得共轭方向 $\boldsymbol{d}_1 = -\boldsymbol{g}_1 + \beta_0\boldsymbol{d}_0$,

精确线搜索 $\boldsymbol{x}_2 = \boldsymbol{x}_1 + \alpha_1\boldsymbol{d}_1$

待定参数 待定参数

令 $\boldsymbol{d}_2 = -\boldsymbol{g}_2 + \beta_0 \boldsymbol{d}_0 + \beta_1 \boldsymbol{d}_1$

选择参数 β_0，β_1，使满足共轭条件

$$\begin{cases} \boldsymbol{d}_2^{\mathrm{T}} \boldsymbol{A} \boldsymbol{d}_0 = 0 \\ \boldsymbol{d}_2^{\mathrm{T}} \boldsymbol{A} \boldsymbol{d}_1 = 0 \end{cases} \Longrightarrow \begin{cases} \beta_0 = \dfrac{\boldsymbol{g}_2^{\mathrm{T}} \boldsymbol{A} \boldsymbol{d}_0}{\boldsymbol{d}_0^{\mathrm{T}} \boldsymbol{A} \boldsymbol{d}_0} = 0 \\ \beta_1 = \dfrac{\boldsymbol{g}_2^{\mathrm{T}} \boldsymbol{A} \boldsymbol{d}_1}{\boldsymbol{d}_1^{\mathrm{T}} \boldsymbol{A} \boldsymbol{d}_1} = \dfrac{\boldsymbol{g}_2^{\mathrm{T}} (\boldsymbol{g}_2 - \boldsymbol{g}_1)}{\boldsymbol{d}_1^{\mathrm{T}} (\boldsymbol{g}_2 - \boldsymbol{g}_1)} = \dfrac{\boldsymbol{g}_2^{\mathrm{T}} \boldsymbol{g}_2}{\boldsymbol{g}_1^{\mathrm{T}} \boldsymbol{g}_1} \end{cases}$$

新的共轭方向 $\boldsymbol{d}_2 = -\boldsymbol{g}_2 \qquad + \beta_1 \boldsymbol{d}_1$

$$\beta_1 = \frac{\boldsymbol{g}_2^{\mathrm{T}} (\boldsymbol{g}_2 - \boldsymbol{g}_1)}{\boldsymbol{d}_1^{\mathrm{T}} (\boldsymbol{g}_2 - \boldsymbol{g}_1)} = \frac{\boldsymbol{g}_2^{\mathrm{T}} \boldsymbol{g}_2}{\boldsymbol{g}_1^{\mathrm{T}} \boldsymbol{g}_1}$$

精确线搜索 $\boldsymbol{x}_3 = \boldsymbol{x}_2 + \alpha_2 \boldsymbol{d}_2$

对第 k 次迭代，令

$$d_k = -g_k + \beta_0 d_0 + \beta_1 d_1 + \beta_2 d_2 + \cdots + \beta_{k-1} d_{k-1}$$

确定参数 $\beta_0, \beta_1, \beta_2, \cdots, \beta_{k-1}$ 使满足共轭条件

$$\begin{cases} d_k^{\mathrm{T}} A d_0 = 0 \\ d_k^{\mathrm{T}} A d_1 = 0 \\ \vdots \\ d_k^{\mathrm{T}} A d_{k-1} = 0 \end{cases}$$

$$\boxed{\begin{aligned} &\beta_i = 0, \quad i = 0, 1, \cdots, k-2, \\ &\beta_{k-1} = \frac{g_k^{\mathrm{T}}(g_k - g_{k-1})}{d_{k-1}^{\mathrm{T}}(g_k - g_{k-1})} = \frac{g_k^{\mathrm{T}} g_k}{g_{k-1}^{\mathrm{T}} g_{k-1}}. \end{aligned}}$$

线性共轭梯度法

$$d_k = -g_k \qquad + \beta_{k-1} d_{k-1}$$

$$\boxed{\begin{aligned} &x_{k+1} = x_k + \alpha_k d_k, \\ &d_k = -g_k + \beta_{k-1} d_{k-1} \\ &\beta_{k-1} = \frac{g_k^{\mathrm{T}} g_k}{g_{k-1}^{\mathrm{T}} g_{k-1}} \end{aligned}}$$

线性共轭梯度法

初始步： 取 $\boldsymbol{d}_{-1}=\boldsymbol{0}$, 初始点 $\boldsymbol{x}_0$,终止参数 $\varepsilon \geqslant 0$. 令

$$\boldsymbol{d}_0=-\boldsymbol{g}_0 \qquad k=0$$

迭代步： 若 $\|\boldsymbol{g}_k\| \leqslant \varepsilon$, 算法终止; 否则，计算

$$\boldsymbol{d}_k=-\boldsymbol{g}_k+\beta_{k-1}\boldsymbol{d}_{k-1}$$ 其中， $$\beta_{k-1}=\frac{\boldsymbol{g}_k^{\mathrm{T}}\boldsymbol{g}_k}{\boldsymbol{g}_{k-1}^{\mathrm{T}}\boldsymbol{g}_{k-1}}$$

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

线性共轭梯度法性质

定理 对严格凸二次函数 $f(\boldsymbol{x}) = \frac{1}{2}\boldsymbol{x}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{x} - \boldsymbol{b}^{\mathrm{T}}\boldsymbol{x}$

最优步长规则下的共轭梯度法经 $m \leqslant n$ 步迭代后终

止, 且对任意的 $0 \leqslant k \leqslant m-1$ ，成立

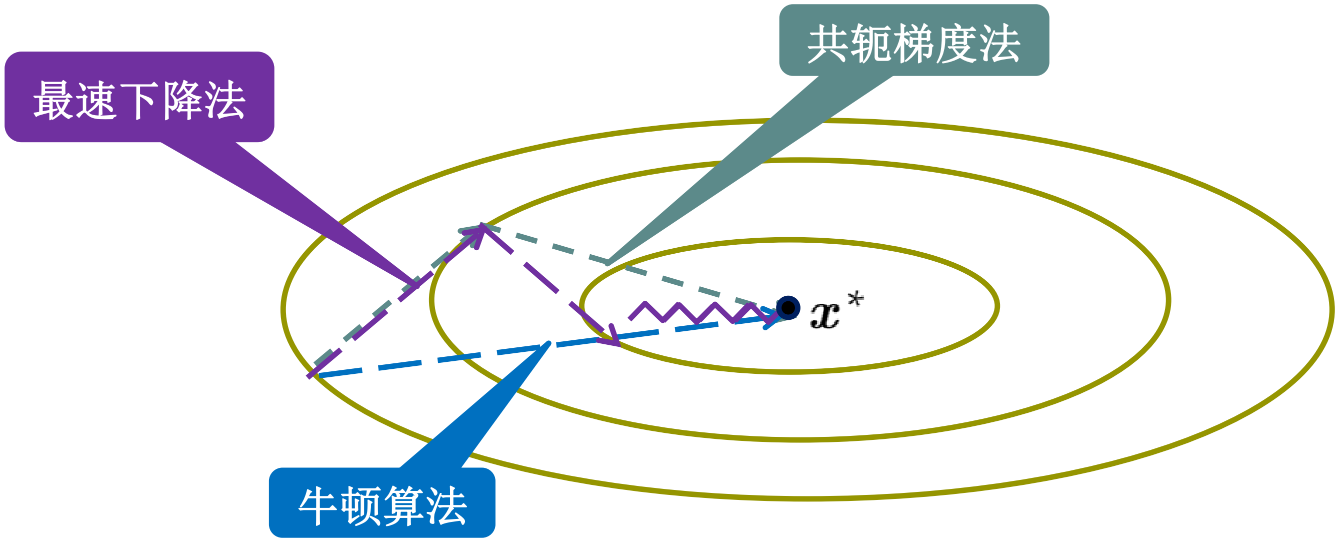
$$\boxed{\boldsymbol{d}_k^{\mathrm{T}}\boldsymbol{A}\boldsymbol{d}_j = 0, \quad \boldsymbol{g}_k^{\mathrm{T}}\boldsymbol{g}_j = \boldsymbol{g}_k^{\mathrm{T}}\boldsymbol{d}_j = 0, \quad j \leqslant k-1,}$$

$$\boxed{\boldsymbol{d}_k^{\mathrm{T}}\boldsymbol{g}_k = -\boldsymbol{g}_k^{\mathrm{T}}\boldsymbol{g}_k,}$$

搜索方向的正交性

搜索方向
的下降性

最速下降法、牛顿法、共轭梯度法的比较



例 $\min x_1^2+\frac{1}{2}x_2^2+\frac{1}{2}x_3^2$

解：取初始点 $\boldsymbol{x}_0=(1;1;1)$ 。迭代过程：

k	$\boldsymbol{x}_k$	g_k	β_{k-1}	d_k	α_k
0	$(1,1,1)^T$	$(2,1,1)^T$	0	$-(2,1,1)^T$	$\frac{3}{5}$
1	$\left(-\frac{1}{5},\frac{2}{5},\frac{2}{5}\right)^T$	$\left(-\frac{2}{5},\frac{2}{5},\frac{2}{5}\right)^T$	$\frac{2}{25}$	$\frac{6}{25}(1,-2,-2)^T$	$\frac{5}{6}$
2	$(0,0,0)^T$				

观察：算法两步迭代后终止。

问题：算法的迭代步数与什么有关？

收敛速度

定理 对严格凸二次函数 $f(x)=\frac{1}{2}\boldsymbol{x}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{x}-\boldsymbol{b}^{\mathrm{T}}\boldsymbol{x}$

若系数矩阵$\boldsymbol{A}$有 r 个相异特征根，则最优步长规则下的共轭梯度法至多 r 步迭代后终止。

定理 对严格凸二次函数 $f(\boldsymbol{x})=\frac{1}{2}\boldsymbol{x}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{x}-\boldsymbol{b}^{\mathrm{T}}\boldsymbol{x}$

最优步长规则下的共轭梯度法产生的迭代点列满足

$$\|\boldsymbol{x}_k-\boldsymbol{x}^*\|_{\boldsymbol{A}}\leqslant\left(\frac{\kappa(\boldsymbol{A})-1}{\kappa(\boldsymbol{A})+1}\right)^k\|\boldsymbol{x}_0-\boldsymbol{x}^*\|_{\boldsymbol{A}}.$$

其中，$\boldsymbol{x}^*$ 为目标函数的最小值点，

矩阵$\boldsymbol{A}$的条件数 $\kappa(\boldsymbol{A})=\|\boldsymbol{A}\|\|\boldsymbol{A}^{-1}\|=\dfrac{\lambda_{\max}(\boldsymbol{A})}{\lambda_{\min}(\boldsymbol{A})}$

线性共轭梯度法小结

对大规模线性方程组问题，线性共轭梯度法计算量和储存量小，快速收敛，计算效率高于传统的Gauss消元法和分解算法.

对于小规模线性方程组问题,Gauss消元法更简便.

三、非线性共轭梯度法

严格凸二次函数 $f(\boldsymbol{x}) = \frac{1}{2}\boldsymbol{x}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{x} - \boldsymbol{b}^{\mathrm{T}}\boldsymbol{x},$

$$\boldsymbol{x}_{k+1} = \boldsymbol{x}_k + \alpha_k \boldsymbol{d}_k, \qquad \boldsymbol{d}_k = -\boldsymbol{g}_k + \beta_{k-1}\boldsymbol{d}_{k-1}.$$

$$\beta_{k-1} = \boxed{\frac{\boldsymbol{g}_k^{\mathrm{T}}(\boldsymbol{g}_k - \boldsymbol{g}_{k-1})}{\boldsymbol{d}_{k-1}^{\mathrm{T}}(\boldsymbol{g}_k - \boldsymbol{g}_{k-1})}} \quad \text{或} \quad \beta_{k-1} = \boxed{\frac{\boldsymbol{g}_k^{\mathrm{T}}\boldsymbol{g}_k}{\boldsymbol{g}_{k-1}^{\mathrm{T}}\boldsymbol{g}_{k-1}}}$$

$$\boxed{\begin{array}{l} \boldsymbol{d}_k^{\mathrm{T}}\boldsymbol{A}\boldsymbol{d}_j = 0, \quad \boldsymbol{g}_k^{\mathrm{T}}\boldsymbol{g}_j = \boldsymbol{g}_k^{\mathrm{T}}\boldsymbol{d}_j = 0, \quad j \leqslant k-1, \\ \boldsymbol{d}_k^{\mathrm{T}}\boldsymbol{g}_k = -\boldsymbol{g}_k^{\mathrm{T}}\boldsymbol{g}_k, \end{array}}$$

线性共轭梯度法的性质

将上述方法推广到非线性优化，便得非线性共轭梯度法。

$$\frac{\boldsymbol{g}_k^{\mathrm{T}}(\boldsymbol{g}_k - \boldsymbol{g}_{k-1})}{\boldsymbol{d}_{k-1}^{\mathrm{T}}(\boldsymbol{g}_k - \boldsymbol{g}_{k-1})} \overset{?}{=} \frac{\boldsymbol{g}_k^{\mathrm{T}}\boldsymbol{g}_k}{\boldsymbol{g}_{k-1}^{\mathrm{T}}\boldsymbol{g}_{k-1}} \qquad \textbf{NO!}$$

不同的 β_{k-1} 取值得到不同的非线性共轭梯度法。

非线性共轭梯度法

搜索方向： $\boldsymbol{d}_k = -\boldsymbol{g}_k + \beta_{k-1}\boldsymbol{d}_{k-1}$

$$\beta_{k-1} = \frac{\boldsymbol{g}_k^{\mathrm{T}}(\boldsymbol{g}_k - \boldsymbol{g}_{k-1})}{\boldsymbol{d}_{k-1}^{\mathrm{T}}(\boldsymbol{g}_k - \boldsymbol{g}_{k-1})}$$

Crowder-Wolfe公式

$$\beta_{k-1} = \frac{\boldsymbol{g}_k^{\mathrm{T}}\boldsymbol{g}_k}{\boldsymbol{g}_{k-1}^{\mathrm{T}}\boldsymbol{g}_{k-1}}$$

Fletcher-Reeves 公式

$$\beta_{k-1} = \frac{\boldsymbol{g}_k^{\mathrm{T}}(\boldsymbol{g}_k - \boldsymbol{g}_{k-1})}{\boldsymbol{g}_{k-1}^{\mathrm{T}}\boldsymbol{g}_{k-1}}$$

Polak-Ribiere公式

$$\beta_{k-1} = -\frac{\boldsymbol{g}_k^{\mathrm{T}}\boldsymbol{g}_k}{\boldsymbol{d}_{k-1}^{\mathrm{T}}\boldsymbol{g}_{k-1}}$$

Dixon公式

$$\beta_{k-1} = \frac{\boldsymbol{g}_k^{\mathrm{T}}\boldsymbol{g}_k}{\boldsymbol{d}_{k-1}^{\mathrm{T}}(\boldsymbol{g}_k - \boldsymbol{g}_{k-1})}$$

Dai-Yuan公式

步长： 精确步长松弛为非精确步长(如Wolfe步长)

数值效果与收敛速度

1. FR方法和PR方法是最常见的共轭梯度法。从数值效果上， PR方法优于FR方法， 但其理论性质不及FR方法.

2. 线性共轭梯度法具有二次终止性. 非线性共轭梯度法无此性质.

3. 非线性共轭梯度法线性收敛(Crowder & Wolfe, 1972). 有例为证(Powell, 1976)

线性收敛的例子 $f(\boldsymbol{x})=\begin{cases}\frac{1}{2}\boldsymbol{x}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{x}, & \text{若 } \boldsymbol{x}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{x}\leqslant 4,\\ \frac{1}{2}\boldsymbol{x}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{x}+(\boldsymbol{b}^{\mathrm{T}}\boldsymbol{x})(\boldsymbol{x}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{x}-4)^2, & \text{其他}.\end{cases}$

$$\boldsymbol{b}=\Big(\frac{\sqrt{6}}{12};\frac{4}{9}\sqrt{\frac{6}{5}};\frac{-41\sqrt{5}}{90}\Big),\quad \boldsymbol{A}=\mathrm{diag}\Big(\frac{1}{10},1,1\Big).$$

全局最优解：$\boldsymbol{x}^*=\mathbf{0}$。以 $\boldsymbol{x}_0=\Big(\frac{5\sqrt{6}}{2};0;\frac{\sqrt{5}}{2}\Big)$ 为初始点，应用最优步长规则下的PR方法,

则第一次迭代后产生的所有迭代点和搜索过程都在椭球 $\{x\in\mathbb{R}^3\mid \boldsymbol{x}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{x}\leqslant 4\}$ 内，且满足 $\|\boldsymbol{x}_k-\boldsymbol{x}^*\|=\sqrt{\frac{105}{6}\Big(\frac{3}{5}\Big)^{k-2}}.$

迭代点列线性收敛到最优值点！

虽然如此，无锯齿现象！

原因分析： 当迭代点进入 $x^{\mathrm{T}}Ax \leqslant 4$ 区域后，尽管目标函数是凸二次函数，但由于初始搜索方向不是负梯度方向，后续产生的搜索方向关于系数矩阵不共轭，算法不有限步终止.

改进举措：迭代若干次后，调用一次负梯度方向(再开始共轭梯度法).

线性共轭梯度法 n步终止。由此得共轭梯度法的 n 步重新开始策略。

Cohen(1972)证明了该共轭梯度法 n 步2阶收敛

$$\|x_{k+n} - x^*\| = O(\|x_k - x^*\|^2).$$

四、共轭梯度法收敛性

搜索方向的下降性：强Wolfe步长规则下的FR方法产生的迭代点列满足 $\boldsymbol{g}_k^{\mathrm{T}}\boldsymbol{d}_k < 0.$

Wolfe步长

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

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

$$0 < \sigma_1 < \sigma_2 < 1$$

定理(FR方法）设目标函数连续可微,梯度在有界水平集上Lipschitz连续.则强Wolfe步长规则下的FR共轭梯度法弱收敛，即满足 $\liminf\limits_{k\to\infty}\|\boldsymbol{g}_k\| = 0.$

- 一般情况下，PR方法没有全局收敛性！

  Powell(1984)给出反例：即使目标函数二次连续可微，水平集有界，并采用最优步长，PR方法也会产生一个包含稳定点的聚点.

- 对一致凸函数，最优步长规则下，PR方法全局收敛。

定理(PR方法) 设目标函数二阶连续可微, 水平集有界. 又设存在常数 $\mu > 0$, 使对任意的 $\boldsymbol{x} \in \mathcal{L}(\boldsymbol{x}_0)$ ，均有

$$\boldsymbol{y}^{\mathrm{T}} \boldsymbol{G}(\boldsymbol{x}) \boldsymbol{y} \geqslant \mu \|\boldsymbol{y}\|^2, \quad \forall\, \boldsymbol{y} \in \mathbb{R}^n.$$

则最优步长规则下的PR方法全局收敛.

最速下降算法、牛顿法、共轭梯度法比较

	初始点选取	计算量与存储	收敛速率	二次终止性
最速下降法	任意	小	线性 有锯齿现象	✕