第4章 逐次逼近法

4.5 共轭梯度法

4.5 共轭梯度法

共轭梯度法（**conjugate gradient method, CG**）是以共轭方向（**conjugate direction**）作为搜索方向的一类算法。

共轭梯度法是由**Hesteness**和**Stiefel**于**1952**年为求解线性方程组而提出的。后来用于求解无约束最优化问题，它是一种重要的数学优化方法。

设线性方程组：

$$\boldsymbol{A}\boldsymbol{x}=\boldsymbol{b} \qquad (4\text{-}62)$$

其中$\boldsymbol{A}$为$\boldsymbol{n}$阶对称正定矩阵，$\boldsymbol{b}\in\mathbf{R}^n$，$\boldsymbol{x}\in\mathbf{R}^n$ 为待求向量。

考察二次函数 $\varphi(\boldsymbol{x}):\mathbf{R}^n\to\mathbf{R}$

$$\varphi(\boldsymbol{x})=\frac{1}{2}(\boldsymbol{A}\boldsymbol{x},\boldsymbol{x})-(\boldsymbol{b},\boldsymbol{x})=\frac{1}{2}\sum_{j=1}^{n}\left(\sum_{i=1}^{n}a_{ij}x_i\right)x_j-\sum_{j=1}^{n}b_jx_j$$

则有

定理4.9 x^*是$Ax=b$的解(A为对称正定矩阵)的充分必要条件

为x^*满足 $$\varphi\left(x^*\right)=\min_{x\in\mathbf{R}^n}\varphi\left(x\right)$$

证 必要性 设$Ax^*=b$, 取$x=x^*+tp$,其中 $t\in\mathbf{R},\ 0\neq p\in\mathbf{R}^n$, 则

$$\varphi\left(x^*+tp\right)=\frac{1}{2}\left(A\left(x^*+tp\right),\ x^*+tp\right)-\left(b,\ x^*+tp\right)$$ 注意到$A=A^T$,

$$=\frac{1}{2}\left(Ax^*,\ x^*\right)+\frac{t}{2}\left(Ax^*,\ p\right)+\frac{t}{2}\left(Ap,\ x^*\right)+\frac{t^2}{2}\left(Ap,\ p\right)-\left(b,\ x^*\right)-t\left(b,\ p\right)$$

$$=\boxed{\frac{1}{2}\left(Ax^*,\ x^*\right)-\left(b,\ x^*\right)}+\boxed{\frac{t}{2}\left[\left(Ax^*,\ p\right)+\left(Ap,\ x^*\right)-2\left(b,\ p\right)\right]}+\frac{t^2}{2}\left(Ap,\ p\right)$$

$$=\varphi\left(x^*\right)+t\left(Ax^*-b,\ p\right)+\frac{t^2}{2}\left(Ap,\ p\right)$$

$$=\varphi\left(x^*\right)+\frac{t^2}{2}\left(Ap,\ p\right)$$

因为A为正定矩阵，所以 $\frac{t^2}{2}\left(Ap,\ p\right)\geq 0$, 故

$$\varphi\left(x^*+tp\right)\geq\varphi\left(x^*\right)$$

即x^*使 $\varphi\left(x\right)$ 达到最小。

$$\varphi\left(\boldsymbol{x}^{*}+t\boldsymbol{p}\right)=\varphi\left(\boldsymbol{x}^{*}\right)+t\left(\boldsymbol{A}\boldsymbol{x}^{*}-\boldsymbol{b},\boldsymbol{p}\right)+\frac{t^{2}}{2}(\boldsymbol{A}\boldsymbol{p},\boldsymbol{p})$$

充分性　设 $\varphi\left(\boldsymbol{x}^{*}\right)=\min_{\boldsymbol{x}\in\mathbf{R}^{n}}\varphi(\boldsymbol{x})$, 则应有 $\varphi\left(\boldsymbol{x}^{*}+t\boldsymbol{p}\right)$ 于t=0取极小值, 也即 $\forall\boldsymbol{p}\in\mathbf{R}^{n}$

$$\left[\varphi'\left(\boldsymbol{x}^{*}+t\boldsymbol{p}\right)\right]\Big|_{t=0}=\left[\left(\boldsymbol{A}\boldsymbol{x}^{*}-\boldsymbol{b},\boldsymbol{p}\right)+t\left(\boldsymbol{A}\boldsymbol{p},\boldsymbol{p}\right)\right]\Big|_{t=0}=\left(\boldsymbol{A}\boldsymbol{x}^{*}-\boldsymbol{b},\boldsymbol{p}\right)=0$$

从而 $\left(\boldsymbol{A}\boldsymbol{x}^{*}-\boldsymbol{b},\boldsymbol{p}\right)=0 \Leftrightarrow$　$\boldsymbol{x}^*$是$\boldsymbol{A}\boldsymbol{x}=\boldsymbol{b}$的解。

又由 $\varphi\left(\boldsymbol{x}^{*}+t\boldsymbol{p}\right)$ 于t=0取极小值的充分条件, 即得

$$\left(\boldsymbol{A}\boldsymbol{p},\boldsymbol{p}\right)=\varphi''(0)>0\quad\forall\boldsymbol{p}\in\mathbf{R}^{n}$$

故$\boldsymbol{A}$必为正定矩阵。

定理4.9说明，求$\boldsymbol{A}\boldsymbol{x}=\boldsymbol{b}$的解的问题等价于求 $\varphi(\boldsymbol{x})$ 的最小值问题。

在介绍共轭梯度法前, 先介绍较为直观的最速下降法。

4.5.1 最速下降法

取初始向量$\boldsymbol{x}_0$，从$\boldsymbol{x}_0$出发构造向量序列$\{\boldsymbol{x}_k\}$使 $\varphi(\boldsymbol{x}_{k-1}) > \varphi(\boldsymbol{x}_k)$

构造方法：选取方向$\boldsymbol{x}_0$，使$\varphi(\boldsymbol{x})$在$\boldsymbol{x}_0$处沿$\boldsymbol{y}_0$方向减小的速度最快。

据多元函数场论可知，$\boldsymbol{y}_0$应为$\varphi(\boldsymbol{x})$的负梯度方向，即

$$\boldsymbol{y}_0 = -\nabla\varphi(\boldsymbol{x}_0) = \boldsymbol{r}_0 = \boldsymbol{b} - \boldsymbol{A}\boldsymbol{x}_0 \qquad \varphi(\boldsymbol{x}) = \frac{1}{2}\boldsymbol{x}^T\boldsymbol{A}\boldsymbol{x} - \boldsymbol{x}^T\boldsymbol{b}$$

再在方向$\boldsymbol{r}_0$上进行一维极小搜索，即在 $\boldsymbol{x}_0 + \alpha\boldsymbol{r}_0$ 中选取 α_0 使得 $\varphi(\boldsymbol{x}_0 + \alpha\boldsymbol{r}_0)$ 极小，即求 $\min\limits_{\alpha\in R}\varphi(\boldsymbol{x}_0 + \alpha\boldsymbol{r}_0)$

$$\frac{\mathrm{d}\varphi(\boldsymbol{x}_0 + \alpha\boldsymbol{r}_0)}{\mathrm{d}\alpha} = \frac{\mathrm{d}}{\mathrm{d}\alpha}\left[\varphi(\boldsymbol{x}_0) + \alpha(\boldsymbol{A}\boldsymbol{x}_0 - \boldsymbol{b}, \boldsymbol{r}_0) + \frac{\alpha^2}{2}(\boldsymbol{A}\boldsymbol{r}_0, \boldsymbol{r}_0)\right]$$

$$= -(\boldsymbol{r}_0, \boldsymbol{r}_0) + \alpha(\boldsymbol{A}\boldsymbol{r}_0, \boldsymbol{r}_0) = 0$$

则 $\alpha=\alpha_0=\dfrac{(\boldsymbol{r}_0,\boldsymbol{r}_0)}{(\boldsymbol{A}\boldsymbol{r}_0,\boldsymbol{r}_0)}$，由于 $\dfrac{\mathrm{d}^2}{\mathrm{d}\alpha^2}\varphi(\boldsymbol{x}_0+\alpha\boldsymbol{r}_0)=(\boldsymbol{A}\boldsymbol{r}_0,\boldsymbol{r}_0)>0$

故 $\min\limits_{\alpha\in R}\varphi(\boldsymbol{x}_0+\alpha\boldsymbol{r}_0)=\varphi(\boldsymbol{x}_0+\alpha\boldsymbol{r}_0)$

令 $\boldsymbol{x}_1=\boldsymbol{x}_0+\alpha\boldsymbol{r}_0$，重复上面的过程，可得

$$\begin{cases}\boldsymbol{r}_k=\boldsymbol{b}-\boldsymbol{A}\boldsymbol{x}_k\\[2ex] \alpha_k=\dfrac{(\boldsymbol{r}_k,\boldsymbol{r}_k)}{(\boldsymbol{A}\boldsymbol{r}_k,\boldsymbol{r}_k)}\\[2ex] \boldsymbol{x}_{k+1}=\boldsymbol{x}_k+\alpha_k\boldsymbol{r}_k\end{cases} \tag{4-63}$$

由于 $\varphi(\boldsymbol{x}_0)>\varphi(\boldsymbol{x}_1)>\cdots>\varphi(\boldsymbol{x}_k)>\cdots\geq\varphi(\boldsymbol{x}^*)$，$\{\varphi(\boldsymbol{x}_k)\}$ 存在极限，

而且 $$\lim_{k\to\infty}\boldsymbol{x}_k=\boldsymbol{x}^*=\boldsymbol{A}^{-1}\boldsymbol{b}$$

还可以证明

（1） $(\boldsymbol{r}_{k+1},\boldsymbol{r}_k)=0\quad k=0,1,2,\cdots$ （4-64）

（2） $\|\boldsymbol{x}_k-\boldsymbol{x}^*\|_A\leq\left(\dfrac{\lambda_1-\lambda_n}{\lambda_1+\lambda_n}\right)^k\|\boldsymbol{x}_0-\boldsymbol{x}^*\|_A$ （4-65）

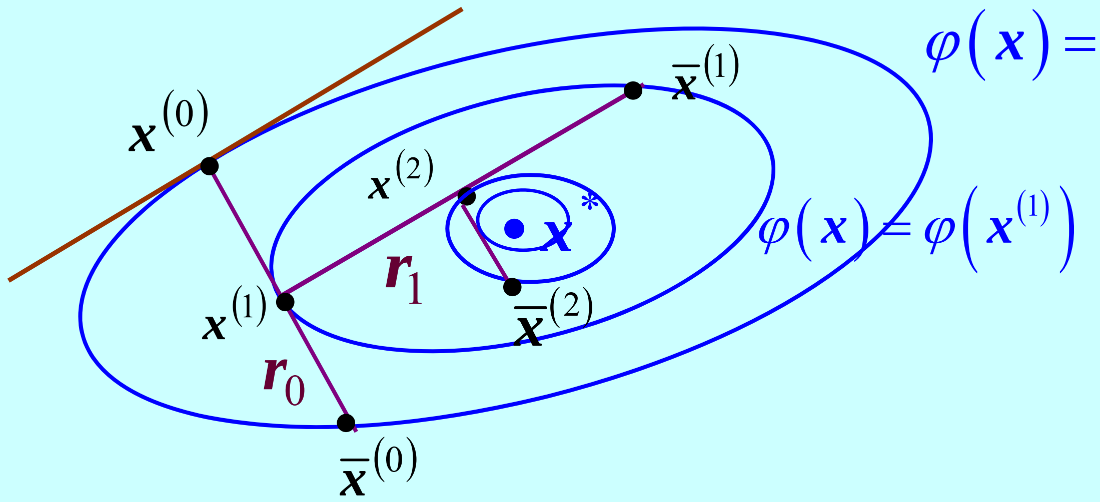
其中 λ_1,λ_n 分别为$\boldsymbol{A}$的模最大和最小特征值，$\|\boldsymbol{x}\|_A=(\boldsymbol{A}\boldsymbol{x},\boldsymbol{x})^{\frac{1}{2}}$。

最速下降的几何意义

取定初始值$\boldsymbol{x}^{(0)}$，于超椭圆 $\varphi(\boldsymbol{x})=\varphi\left(\boldsymbol{x}^{(0)}\right)$ 上过$\boldsymbol{x}^{(0)}$，作其内法线$\boldsymbol{r}^{(0)}$，它就是 $\varphi(\boldsymbol{x})$ 在$\boldsymbol{x}^{(0)}$点的最速下降方向，然后在射线 $\boldsymbol{x}^{(0)}+\alpha\boldsymbol{r}^{(0)}$ 上求与 $\varphi(\boldsymbol{x})=\varphi\left(\boldsymbol{x}^{(0)}\right)$ 的另一交点 $\overline{\boldsymbol{x}}^{(0)}$，$\overline{\boldsymbol{x}}^{(0)}$与 $\boldsymbol{x}^{(0)}$ 的中点即为$\boldsymbol{x}^{(1)}$，它也是 $\boldsymbol{x}^{(0)}+\alpha\boldsymbol{r}^{(0)}$ 与 $\varphi(\boldsymbol{x})=\varphi\left(\boldsymbol{x}^{(1)}\right)$ 的切点。

过$\boldsymbol{x}^{(1)}$再作 $\varphi(\boldsymbol{x})=\varphi\left(\boldsymbol{x}^{(1)}\right)$ 的内法线$\boldsymbol{r}^{(1)}$，然后沿着此方向求 $\overline{\boldsymbol{x}}^{(1)}$，$\boldsymbol{x}^{(2)}$，等等…，如此继续下去就得到最速下降的迭代序列。

$$\varphi(\boldsymbol{x})=\frac{1}{2}\boldsymbol{x}^T\boldsymbol{A}\boldsymbol{x}-\boldsymbol{x}^T\boldsymbol{b}$$



验证 $(\boldsymbol{r}_{k+1}, \boldsymbol{r}_k) = 0 \quad k = 0, 1, 2, \cdots$ （4-64）

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

得 $\boldsymbol{r}_{k+1} = \boldsymbol{b} - \boldsymbol{A}\boldsymbol{x}_{k+1} = \boldsymbol{b} - \boldsymbol{A}(\boldsymbol{x}_k + \alpha_k \boldsymbol{r}_k) = \boldsymbol{r}_k - \alpha_k \boldsymbol{A}\boldsymbol{r}_k$

从而 $(\boldsymbol{r}_{k+1}, \boldsymbol{r}_k) = (\boldsymbol{r}_k, \boldsymbol{r}_k) - \alpha_k (\boldsymbol{A}\boldsymbol{r}_k, \boldsymbol{r}_k)$ $\qquad \alpha_k = \dfrac{(\boldsymbol{r}_k, \boldsymbol{r}_k)}{(\boldsymbol{A}\boldsymbol{r}_k, \boldsymbol{r}_k)}$

$$= (\boldsymbol{r}_k, \boldsymbol{r}_k) - \frac{(\boldsymbol{r}_k, \boldsymbol{r}_k)}{\cancel{(\boldsymbol{A}\boldsymbol{r}_k, \boldsymbol{r}_k)}} \cancel{(\boldsymbol{A}\boldsymbol{r}_k, \boldsymbol{r}_k)}$$

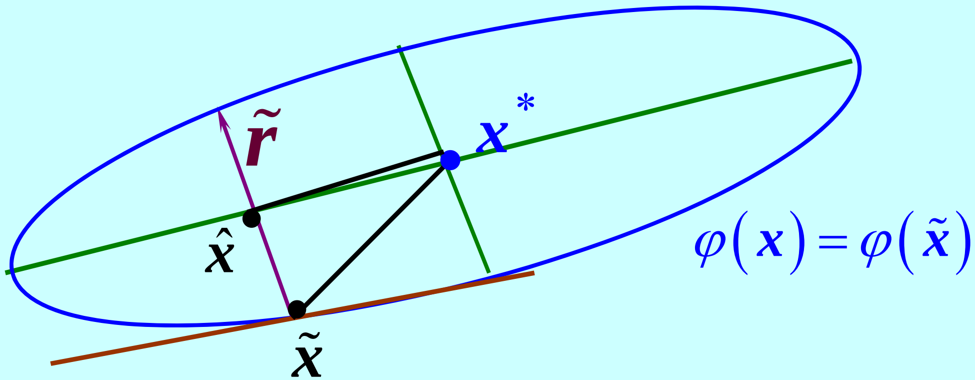
$$= (\boldsymbol{r}_k, \boldsymbol{r}_k) - (\boldsymbol{r}_k, \boldsymbol{r}_k)$$

$$= 0$$

当 λ_1 与 λ_n 相差很大时，据（4-65）可知, $\{x_k\}$ 收敛很慢，而且当 $\|r_k\|$ 很小时,由于舍入误差的影响,由（4-63）的计算将出现不稳定现象, 所以在实际计算中很少使用最速下降法。

这是因为当 λ_1 与 λ_n 相差很大时，椭球面变得非常扁平；如果 $\tilde{x}$ 位于它的较平坦的一面时, 其内法线 $\tilde{r}$ 就与 $\hat{x}$ 和椭球中心 x^* 的连线方向几乎垂直，这样 $\hat{x}$ 与 x^* 之间的距离就差不多等于 $\tilde{x}$ 与 x^* 之间的距离，从而使得最速下降法的收敛速度变得非常慢。

$$\|x_k - x^*\|_A \le \left(\frac{\lambda_1 - \lambda_n}{\lambda_1 + \lambda_n}\right)^k \|x_0 - x^*\|_A$$



从上述几何解释说明, 负梯度方向并非最合适的方向, 有时它与目标的偏差太大。能否选出比负梯度方向更好而计算量又不是很大的方向为新的下降方向呢？答案是肯定的。

$$\varphi(\boldsymbol{x})=\frac{1}{2}\boldsymbol{x}^T\boldsymbol{A}\boldsymbol{x}-\boldsymbol{x}^T\boldsymbol{b} \quad \boldsymbol{r}_k=\boldsymbol{b}-\boldsymbol{A}\boldsymbol{x}_k \quad \alpha_k=\frac{(\boldsymbol{r}_k,\boldsymbol{r}_k)}{(\boldsymbol{A}\boldsymbol{r}_k,\boldsymbol{r}_k)} \quad \boldsymbol{x}_{k+1}=\boldsymbol{x}_k+\alpha_k\boldsymbol{r}_k$$

4.5.2 共轭梯度法（简称CG法）

$$\left\|\boldsymbol{x}_k-\boldsymbol{x}^*\right\|_A\le\left(\frac{\lambda_1-\lambda_n}{\lambda_1+\lambda_n}\right)^k\left\|\boldsymbol{x}_0-\boldsymbol{x}^*\right\|_A$$

从上式可以看出，我们采用一维搜索所沿的方向$\boldsymbol{r}_0, \boldsymbol{r}_1, \ldots, \boldsymbol{r}_k$可能使的收敛速度缓慢, 因此, 我们需要另找一组方向 $\boldsymbol{p}_0, \boldsymbol{p}_1, \cdots, \boldsymbol{p}_k$进行一维极小搜索。设按方向 $\boldsymbol{p}_0, \boldsymbol{p}_1, \cdots, \boldsymbol{p}_k$已经进行$k$次一维搜索, 已求出$\boldsymbol{x}_k$, 下一步确定进行求解极小问题

$$\min\varphi(\boldsymbol{x}_k+\alpha\boldsymbol{p}_k)$$

与最速下降法中的方法一样, 令

$$\frac{\mathrm{d}\varphi(\boldsymbol{x}_k+\alpha\boldsymbol{p}_k)}{\mathrm{d}\alpha}=\frac{\mathrm{d}}{\mathrm{d}\alpha}\left[\varphi(\boldsymbol{x}_k)+\alpha(\boldsymbol{A}\boldsymbol{x}_k-\boldsymbol{b},\boldsymbol{p}_k)+\frac{\alpha^2}{2}(\boldsymbol{A}\boldsymbol{p}_k,\boldsymbol{p}_k)\right]$$

$$=(\boldsymbol{A}\boldsymbol{x}_k-\boldsymbol{b},\boldsymbol{p}_k)+\alpha(\boldsymbol{A}\boldsymbol{p}_k,\boldsymbol{p}_k)$$

$$=-(\boldsymbol{r}_k,\boldsymbol{p}_k)+\alpha(\boldsymbol{A}\boldsymbol{p}_k,\boldsymbol{p}_k)=0$$

$$-(\boldsymbol{r}_k,\boldsymbol{p}_k)+\alpha(\boldsymbol{A}\boldsymbol{p}_k,\boldsymbol{p}_k)=0$$

解得
$$\alpha=\alpha_k=\frac{(\boldsymbol{r}_k,\boldsymbol{p}_k)}{(\boldsymbol{A}\boldsymbol{p}_k,\boldsymbol{p}_k)} \qquad \text{（4-66）}$$

从而得到下一个近似解和对应的负梯度向量

$$\boldsymbol{x}_{k+1}=\boldsymbol{x}_k+\alpha_k\boldsymbol{p}_k \qquad \text{（4-67）}$$

$$\boldsymbol{r}_{k+1}=\boldsymbol{b}-\boldsymbol{A}\boldsymbol{x}_{k+1}=\boldsymbol{b}-\boldsymbol{A}(\boldsymbol{x}_k+\alpha_k\boldsymbol{p}_k)$$

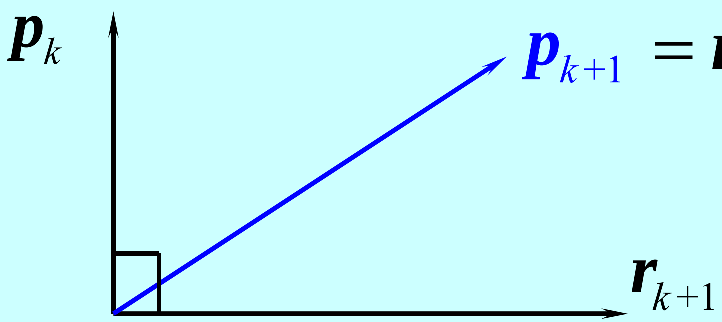
$$=\boldsymbol{b}-\boldsymbol{A}\boldsymbol{x}_k-\alpha_k\boldsymbol{A}\boldsymbol{p}_k=\boldsymbol{r}_k-\alpha_k\boldsymbol{A}\boldsymbol{p}_k \qquad \text{（4-68）}$$

下面讨论 $\boldsymbol{p}_0,\boldsymbol{p}_1,\cdots,\boldsymbol{p}_k\cdots$的取法。

开始时取 $\boldsymbol{p}_0=\boldsymbol{r}_0$, 如果$k\geq 1$时， 我们可以证明这样的事实：

上一步$\boldsymbol{x}_k$的下降方向$\boldsymbol{p}_k$一定与$\boldsymbol{r}_{k+1}$(当前步$\boldsymbol{x}_{k+1}$的负梯度方向)是正交的。

$$\boldsymbol{p}_{k+1}=\boldsymbol{r}_{k+1}+\beta\boldsymbol{p}_k$$



$$\min \varphi(\boldsymbol{x}_k + \alpha \boldsymbol{p}_k) \qquad \alpha = \alpha_k = \frac{(\boldsymbol{r}_k, \boldsymbol{p}_k)}{(\boldsymbol{A}\boldsymbol{p}_k, \boldsymbol{p}_k)} \quad (4\text{-}66) \qquad \boldsymbol{x}_{k+1} = \boldsymbol{x}_k + \alpha_k \boldsymbol{p}_k$$

$$\boldsymbol{r}_{k+1} = \boldsymbol{r}_k - \alpha_k \boldsymbol{A}\boldsymbol{p}_k \quad (4\text{-}68) \qquad \boldsymbol{p}_{k+1} = \boldsymbol{r}_{k+1} + \beta_k \boldsymbol{p}_k \quad (4\text{-}72)$$

$$\begin{aligned}&\min_{\beta} \varphi\left(\boldsymbol{x}_{k+1} + \alpha_k (\boldsymbol{r}_{k+1} + \beta \boldsymbol{p}_k)\right) \\ &= \min_{\beta} \varphi\left((\boldsymbol{x}_{k+1} + \alpha_k \boldsymbol{r}_{k+1}) + \alpha_k \beta \boldsymbol{p}_k\right)\end{aligned}$$

令 $$\frac{\mathrm{d}\varphi\left((\boldsymbol{x}_{k+1} + \alpha_k \boldsymbol{r}_{k+1}) + \alpha_k \beta \boldsymbol{p}_k\right)}{\mathrm{d}\beta}$$

$$= \frac{\mathrm{d}}{\mathrm{d}\beta}\left[\varphi(\boldsymbol{x}_{k+1} + \alpha_k \boldsymbol{r}_{k+1}) + \alpha_k \beta\left(\boldsymbol{A}(\boldsymbol{x}_{k+1} + \alpha_k \boldsymbol{r}_{k+1}) - \boldsymbol{b}, \boldsymbol{p}_k\right) + \frac{(\alpha_k \beta)^2}{2}(\boldsymbol{A}\boldsymbol{p}_k, \boldsymbol{p}_k)\right]$$

$$= \alpha_k (\boldsymbol{A}\boldsymbol{x}_{k+1} - \boldsymbol{b}, \boldsymbol{p}_k) + \alpha_k^2 (\boldsymbol{A}\boldsymbol{r}_{k+1}, \boldsymbol{p}_k) + \alpha_k^2 \beta (\boldsymbol{A}\boldsymbol{p}_k, \boldsymbol{p}_k)$$

$$= -\alpha_k (\boldsymbol{r}_{k+1}, \boldsymbol{p}_k) + \alpha_k^2 (\boldsymbol{A}\boldsymbol{r}_{k+1}, \boldsymbol{p}_k) + \alpha_k^2 \beta (\boldsymbol{A}\boldsymbol{p}_k, \boldsymbol{p}_k) = 0$$

则得 $$\beta_k = -\frac{(\boldsymbol{r}_{k+1}, \boldsymbol{A}\boldsymbol{p}_k)}{(\boldsymbol{p}_k, \boldsymbol{A}\boldsymbol{p}_k)}$$

$$\min \varphi\left(\boldsymbol{x}_k+\alpha \boldsymbol{p}_k\right) \qquad \alpha=\alpha_k=\frac{\left(\boldsymbol{r}_k, \boldsymbol{p}_k\right)}{\left(\boldsymbol{A} \boldsymbol{p}_k, \boldsymbol{p}_k\right)} \quad \text{（4-66）} \qquad \boldsymbol{x}_{k+1}=\boldsymbol{x}_k+\alpha_k \boldsymbol{p}_k$$

$$\boldsymbol{r}_{k+1}=\boldsymbol{r}_k-\alpha_k \boldsymbol{A} \boldsymbol{p}_k \quad \text{（4-68）}$$

$$\boldsymbol{p}_{k+1}=\boldsymbol{r}_{k+1}+\beta_k \boldsymbol{p}_k \quad \text{（4-72）} \qquad \beta_k=-\frac{\left(\boldsymbol{r}_{k+1}, \boldsymbol{A} \boldsymbol{p}_k\right)}{\left(\boldsymbol{p}_k, \boldsymbol{A} \boldsymbol{p}_k\right)} \quad \text{（4-73）}$$

由(4-72)、(4-73)得到的$\boldsymbol{p}_k$，还可用于简化α_k的计算。

因为由(4-68)与(4-66)有

$$\begin{aligned}
\left(\boldsymbol{r}_{k+1}, \boldsymbol{p}_k\right) &= \left(\boldsymbol{r}_k-\alpha_k \boldsymbol{A} \boldsymbol{p}_k, \boldsymbol{p}_k\right)=\left(\boldsymbol{r}_k, \boldsymbol{p}_k\right)-\alpha_k\left(\boldsymbol{A} \boldsymbol{p}_k, \boldsymbol{p}_k\right) \\
&=\left(\boldsymbol{r}_k, \boldsymbol{p}_k\right)-\frac{\left(\boldsymbol{r}_k, \boldsymbol{p}_k\right)}{\cancel{\left(\boldsymbol{A} \boldsymbol{p}_k, \boldsymbol{p}_k\right)}} \cancel{\left(\boldsymbol{A} \boldsymbol{p}_k, \boldsymbol{p}_k\right)} \\
&=\left(\boldsymbol{r}_k, \boldsymbol{p}_k\right)-\left(\boldsymbol{r}_k, \boldsymbol{p}_k\right)=0
\end{aligned}$$

$$\alpha_k=\frac{(\boldsymbol{r}_k,\boldsymbol{p}_k)}{(\boldsymbol{A}\boldsymbol{p}_k,\boldsymbol{p}_k)}\qquad (\boldsymbol{r}_{k+1},\boldsymbol{p}_k)=0\qquad \boldsymbol{p}_k=\boldsymbol{r}_k+\beta_{k-1}\boldsymbol{p}_{k-1}$$

$$(\boldsymbol{r}_k,\boldsymbol{p}_k)=(\boldsymbol{r}_k,\boldsymbol{r}_k+\beta_{k-1}\boldsymbol{p}_{k-1})$$

$$=(\boldsymbol{r}_k,\boldsymbol{r}_k)+\beta_{k-1}(\boldsymbol{r}_k,\boldsymbol{p}_{k-1})=(\boldsymbol{r}_k,\boldsymbol{r}_k)\qquad \text{(4-74)}$$

故有

$$\alpha_k=\frac{(\boldsymbol{r}_k,\boldsymbol{p}_k)}{(\boldsymbol{A}\boldsymbol{p}_k,\boldsymbol{p}_k)}=\frac{(\boldsymbol{r}_k,\boldsymbol{r}_k)}{(\boldsymbol{A}\boldsymbol{p}_k,\boldsymbol{p}_k)}\qquad \text{(4-75)}$$

当$\boldsymbol{r}_k\neq\boldsymbol{0}$时，$\alpha_k>0$。

用前面的已有公式，还可以证明：

（1）$$(\boldsymbol{r}_i,\boldsymbol{r}_j)=0\quad (i\neq j)\qquad \text{(4-76)}$$

（2）$$(\boldsymbol{A}\boldsymbol{p}_i,\boldsymbol{p}_j)=(\boldsymbol{p}_i,\boldsymbol{A}\boldsymbol{p}_j)=0\quad (i\neq j)\qquad \text{(4-77)}$$

即$\boldsymbol{r}_0,\boldsymbol{r}_1,\cdots$构成正交向量组，$\boldsymbol{p}_0,\boldsymbol{p}_1,\cdots$构成$\boldsymbol{A}$-正交向量组。

$$\boldsymbol{r}_{k+1} = \boldsymbol{r}_k - \alpha_k \boldsymbol{A}\boldsymbol{p}_k \qquad \text{（4-68）}$$

$$\left(\boldsymbol{r}_i, \boldsymbol{r}_j\right) = 0 \quad \left(i \neq j\right) \qquad \text{（4-76）}$$

$$\alpha_k = \frac{\left(\boldsymbol{r}_k, \boldsymbol{r}_k\right)}{\left(\boldsymbol{A}\boldsymbol{p}_k, \boldsymbol{p}_k\right)} \qquad \text{（4-75）}$$

利用（4-68）和（4-76）还可简化 β_k 的计算

$$\beta_k = -\frac{\left(\boldsymbol{r}_{k+1}, \boldsymbol{A}\boldsymbol{p}_k\right)}{\left(\boldsymbol{p}_k, \boldsymbol{A}\boldsymbol{p}_k\right)} = -\frac{\left(\boldsymbol{r}_{k+1}, \alpha_k^{-1}\left(\boldsymbol{r}_k - \boldsymbol{r}_{k+1}\right)\right)}{\left(\boldsymbol{p}_k, \boldsymbol{A}\boldsymbol{p}_k\right)}$$

$$\alpha_k \boldsymbol{A}\boldsymbol{p}_k = \boldsymbol{r}_k - \boldsymbol{r}_{k+1}$$

$$= -\frac{\alpha_k^{-1}\left(\boldsymbol{r}_{k+1}, \boldsymbol{r}_k\right)}{\left(\boldsymbol{p}_k, \boldsymbol{A}\boldsymbol{p}_k\right)} + \frac{\alpha_k^{-1}\left(\boldsymbol{r}_{k+1}, \boldsymbol{r}_{k+1}\right)}{\left(\boldsymbol{p}_k, \boldsymbol{A}\boldsymbol{p}_k\right)}$$

$$\alpha_k^{-1} = \frac{\left(\boldsymbol{A}\boldsymbol{p}_k, \boldsymbol{p}_k\right)}{\left(\boldsymbol{r}_k, \boldsymbol{r}_k\right)},$$

$$= \frac{\cancel{\left(\boldsymbol{A}\boldsymbol{p}_k, \boldsymbol{p}_k\right)}}{\left(\boldsymbol{r}_k, \boldsymbol{r}_k\right)} \cdot \frac{\left(\boldsymbol{r}_{k+1}, \boldsymbol{r}_{k+1}\right)}{\cancel{\left(\boldsymbol{A}\boldsymbol{p}_k, \boldsymbol{p}_k\right)}} = \frac{\left(\boldsymbol{r}_{k+1}, \boldsymbol{r}_{k+1}\right)}{\left(\boldsymbol{r}_k, \boldsymbol{r}_k\right)} \qquad \text{（4-78）}$$

当 $\boldsymbol{r}_{k+1} \neq 0$ 时，$\beta_k > 0$。

$$\alpha_k = \frac{\left(\boldsymbol{r}_k, \boldsymbol{r}_k\right)}{\left(\boldsymbol{A}\boldsymbol{p}_k, \boldsymbol{p}_k\right)}, \qquad \beta_k = \frac{\left(\boldsymbol{r}_{k+1}, \boldsymbol{r}_{k+1}\right)}{\left(\boldsymbol{r}_k, \boldsymbol{r}_k\right)}$$

共轭梯度法（CG法）：

（1）任取 $\boldsymbol{x}_0 \in \mathbf{R}^n$

（2）$\boldsymbol{r}_0=\boldsymbol{b}-\boldsymbol{A}\boldsymbol{x}_0$，取 $\boldsymbol{p}_0=\boldsymbol{r}_0$

（3）对 k=0, 1, 2, ⋯

$$\alpha_k = \frac{(\boldsymbol{r}_k,\boldsymbol{r}_k)}{(\boldsymbol{p}_k,\boldsymbol{A}\boldsymbol{p}_k)}$$

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

$$\boldsymbol{r}_{k+1} = \boldsymbol{r}_k - \alpha_k \boldsymbol{A}\boldsymbol{p}_k$$

$$\beta_k = \frac{(\boldsymbol{r}_{k+1},\boldsymbol{r}_{k+1})}{(\boldsymbol{r}_k,\boldsymbol{r}_k)}$$

$$\boldsymbol{p}_{k+1} = \boldsymbol{r}_{k+1} + \beta_k \boldsymbol{p}_k$$

若有$\boldsymbol{r}_k=0$，则$\boldsymbol{r}_k=\boldsymbol{b}-\boldsymbol{A}\boldsymbol{x}_k$。

若有$(\boldsymbol{A}\boldsymbol{p}_k,\boldsymbol{p}_k)=0$时，由于$\boldsymbol{A}$为正定阵，必有$\boldsymbol{p}_k=0$，则由（4-74）有

$$(\boldsymbol{r}_k,\boldsymbol{r}_k)=(\boldsymbol{p}_k,\boldsymbol{r}_k)=0$$

即 $\boldsymbol{r}_k=\boldsymbol{b}-\boldsymbol{A}\boldsymbol{x}_k=\boldsymbol{0}$。

在计算过程中若有$\boldsymbol{r}_k=0$或$(\boldsymbol{A}\boldsymbol{p}_k,\boldsymbol{p}_k)=0$时计算终止，即有$\boldsymbol{x}_k=\boldsymbol{x}^*$。

由于$\boldsymbol{n}$维空间中正交向量组的向量个数最多只有$\boldsymbol{n}$个， 所以$\boldsymbol{r}_0, \boldsymbol{r}_1, \cdots, \boldsymbol{r}_{\boldsymbol{n}}$中至少有一个为零向量，若$\boldsymbol{r}_j=\mathbf{0}$, 则$\boldsymbol{r}_j=\boldsymbol{x}^*$。所以使用**CG**法求解$\boldsymbol{n}$阶线性方程组，理论上最多$\boldsymbol{n}$步便可得到精确解, 因此，也可称为直接法。 但是，由于舍入误差的影响，$\{\boldsymbol{r}_k\}$的正交性很难达到, 所以在实际计算时往往不能在$\boldsymbol{n}$步得到精确解, 因此，通常将**CG**法还是作为逐次逼近法使用。

例1 用CG法解线性方程组 $\begin{pmatrix} 2 & 0 & 1 \\ 0 & 2 & 1 \\ 1 & 1 & 2 \end{pmatrix}\begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}$ $\quad \boldsymbol{x}^* = \begin{pmatrix} 0.5 \\ 0.5 \\ 0 \end{pmatrix}$。

解 显然方程组的系数阵为对称正定阵，取 $\boldsymbol{x}_0 = (0,0,0)^T$

$$\boldsymbol{r}_0 = \boldsymbol{b} - \boldsymbol{A}\boldsymbol{x}_0 = \boldsymbol{b} = \boldsymbol{p}_0 = (1,1,1)^T,$$

对 k=0, 1, 2, …计算：α_k , $\boldsymbol{x}_{k+1}$, $\boldsymbol{r}_{k+1}$, β_k , $\boldsymbol{p}_{k+1}$ 。 $\alpha_0 = \dfrac{(\boldsymbol{r}_0, \boldsymbol{r}_0)}{(\boldsymbol{p}_0, \boldsymbol{A}\boldsymbol{p}_0)} = \dfrac{3}{10}$,

$$\boldsymbol{x}_1 = \boldsymbol{x}_0 + \alpha_0 \boldsymbol{p}_0 = \left(\frac{3}{10}, \frac{3}{10}, \frac{3}{10}\right)^T, \quad \boldsymbol{r}_1 = \boldsymbol{r}_0 - \alpha_0 \boldsymbol{A}\boldsymbol{p}_0 = (1,1,1)^T - \left(\frac{9}{10}, \frac{9}{10}, \frac{12}{10}\right)^T = \left(\frac{1}{10}, \frac{1}{10}, -\frac{2}{10}\right)^T$$

$$\beta_0 = \frac{(\boldsymbol{r}_1, \boldsymbol{r}_1)}{(\boldsymbol{r}_0, \boldsymbol{r}_0)} = \frac{6}{100} \times \frac{1}{3} = \frac{1}{50} \qquad \boldsymbol{p}_1 = r_1 + \beta_0 \boldsymbol{p}_0 = \left(\frac{1}{10}, \frac{1}{10}, -\frac{2}{10}\right)^T + \left(\frac{1}{50}, \frac{1}{50}, \frac{1}{50}\right)^T = \frac{3}{50}(2, 2, -3)^T$$

$$\alpha_1 = \frac{(\boldsymbol{r}_1, \boldsymbol{r}_1)}{(\boldsymbol{p}_1, \boldsymbol{A}\boldsymbol{p}_1)} = \frac{6}{100} \times \frac{2500}{90} = \frac{5}{3}$$

$$\boldsymbol{x}_2 = \boldsymbol{x}_1 + \alpha_1 \boldsymbol{p}_1 = \frac{3}{10}(1,1,1)^T + \frac{1}{10}(2,2,-3)^T = \frac{1}{2}(1,1,0)^T$$

$$\boldsymbol{r}_2 = \boldsymbol{r}_1 - \alpha_1 \boldsymbol{A}\boldsymbol{p}_1 = \frac{1}{10}(1,1,-2)^T - \frac{5}{3} \times \frac{3}{50}(1,1,-2)^T = (0,0,0)^T$$

则方程组的解为 $\boldsymbol{x}_2 = (0.5, 0.5, 0)^T$ 。

对于病态的线性方程组，**CG**法的收敛速度是很慢的，为了改进收敛速度，可以对方程组进行预处理，使系数矩阵的条件数降低，这种方法称预处理共轭梯度法（**PCG**法），**PCG**方法是目前求解病态线性方程组的有效方法之一，因此，它已成为很多计算工作者关心的方法之一，对此有兴趣的读者可以参阅文献**[6]**。

THE END