第二章重点（线性方程组的迭代求解）

三个迭代方法

- **雅各比(Jacobi)方法的计算过程**
  - D: The **main diagonal主对角线** of A
  - L: The **lower triangle下三角阵** of A
  - U: The **upper triangle上三角阵** of A

$$
\begin{aligned}
Ax &= b \\
(D+L+U)x &= b \\
Dx &= b-(L+U)x \\
Dx_{k+1} &= b-(L+U)x_k \\
x_{k+1} &= D^{-1}(b-(L+U)x_k)
\end{aligned}
$$

$$
\begin{aligned}
x_0 &= \text{Initial Vector} \\
x_{k+1} &= D^{-1}(b-(L+U)x_k) \\
x_{k+1,i} &= \frac{1}{a_{ii}}(b_i - \sum_{j\neq i} a_{ij}x_{k,j}) \textbf{ for } 1\le i\le n
\end{aligned}
$$

- **雅各比方法的收敛性证明**，严格对角占优定理，雅各比和高斯赛得收敛性定理的证明需要记（需要前置的谱半径定理，这个定理只需要记住不需要证明，$\rho(A)<1$代表迭代方法可收敛）
  - **严格对角占优定理**:
    - The $n\times n$ matrix $A=(a_{ij})$ is **strictly diagonally dominant**, if for each $1\le i\le n$, $|a_{ii}|>\sum_{j\neq i}|a_{ij}|$
  - **谱半径收敛定理**:
    - 谱半径定义：Let A be an $n\times n$ matrix and $\lambda_1,\dots,\lambda_n$ be the **eigenvalues** of A. The **special radius** $\rho(A)$ is defined as $\max\{|\lambda_1|,\dots,|\lambda_n|\}$
      - 可以定义为$\rho(A)=\max|\lambda_i|$
      - 也可以定义为$\rho(A)=\max\frac{||Ax||_2}{||x||_2}$
    - 谱半径收敛定理即为：**For any initial vector x_0, the iteration $x_{k+1}=Ax_k+b$ converges, where A is an $n\times n$ martix with spectral radius $\rho(A)<1$, i.e., there exists a unique x_* s.t. $\lim_{k\to\infty}x_k=x_*$ and $x_*=Ax_*+b$.**
      - 就是把计算方法化为$x_{k+1}=Ax_k+b$这样的表达式，证明此时的A矩阵的$\rho(A)<1$
  - **雅各比方法的收敛性证明**:
    - 将雅各比方法的表达式化为$x_{k+1}=Ax_k+b$的形式，为$x_{k+1}=-D^{-1}(L+U)x_k+D^{-1}b$
    - 因此我们需要证明$\rho(D^{-1}(L+U))<1$
    - 令λ为$D^{-1}(L+U)$的任意特征值，其对应的特征向量为v
    - 记m为最大**特征值分量**的下标，它使得$|v_m|\ge|v_i|$ $\textbf{for } 1\le i\neq m\le n$
    - 因此我们可以将特征值与特征向量的等式做如下转换：
      - $D^{-1}(L+U)v=\lambda v\Rightarrow(L+U)v=\lambda Dv$
    - 对于上面提到的第m个特征值分量，取该向量第m个分量的绝对值，得到式子如下：

$$
\begin{aligned}
|\lambda||v_m||a_{mm}| &= |\lambda a_{mm}v_m| \\
&= |\sum_{i\neq m} a_{mi}v_i| \\
&\le |v_m|\sum_{i\neq m}|a_{mi}| \\
&< |v_m||a_{mm}|
\end{aligned}
$$

    - 因此得到$|\lambda|<1$，因为λ是一个任意的特征值，所以$\rho(D^{-1}(L+U))<1$得证
- **高斯赛得(Gauss-Seidel)方法的计算过程**
  - D: The main diagonal of A
  - L: The lower triangle of A
  - U: The upper triangle of A

$$
\begin{aligned}
Ax &= b \\
(D+L+U)x &= b \\
(D+L)x &= b-Ux \\
(D+L)x_{k+1} &= b-Ux_k
\end{aligned}
$$

For computation: $x_{k+1} = D^{-1}(b - Ux_k - Lx_{k+1})$
For the proof of convergence: $x_{k+1} = (D+L)^{-1}(b - Ux_k)$

$$\begin{aligned} x_0 &= \text{Initial Vector} \\ x_{k+1} &= D^{-1}(b - Ux_k - Lx_{k+1}) \\ x_{k+1,i} &= \frac{1}{a_{ii}}(b_i - \sum_{j>i} a_{ij}x_{k,j} - \sum_{j<i} a_{ij}x_{k+1,j}) \textbf{ for } 1 \le i \le n \end{aligned}$$

- **高斯赛得方法的收敛性证明**
  - 将高斯赛得方法的表达式化为$x_{k+1} = Ax_k + b$的形式，为$x_{k+1} = -(L+D)^{-1}Ux_k + (L+D)^{-1}b$
  - 因此我们需要证明$\rho((L+D)^{-1}U) < 1$
  - 令λ为$(L+D)^{-1}U$的任意特征值，其对应的特征向量为v
  - 记m为最大**特征值分量**的下标，它使得$|v_m| \ge |v_i|$ $\textbf{for } 1 \le i \ne m \le n$
  - 因此我们可以将特征值与特征向量的等式做如下转换：
    - $(L+D)^{-1}Uv = \lambda v \Rightarrow Uv = \lambda(D+L)v$
  - 对于上面提到的第m个特征值分量，取该向量第m个分量的绝对值，得到式子如下：

$$\begin{aligned} |\lambda||v_m| \cdot \sum_{i>m} |a_{mi}| &< |\lambda||v_m| \cdot (|a_{mm}| - \sum_{i<m} |a_{mi}|) \\ &\le |\lambda| \cdot (|a_{mm}v_m| - \sum_{i<m} |a_{mi}v_i|) \\ &\le |\lambda| \cdot |a_{mm}v_m + \sum_{i<m} a_{mi}v_i| \\ &= |\sum_{i>m} a_{mi}v_i| \\ &\le |v_m| \sum_{i>m} |a_{mi}| \end{aligned}$$

  - 因此得到$|\lambda| < 1$，因为λ是一个任意的特征值，所以$\rho((L+D)^{-1}U) < 1$得证
- 过松弛(SOR)方法稍微了解即可
  - D: The main diagonal of A
  - L: The lower triangle of A
  - U: The upper triangle of A

$$\begin{aligned} Ax &= b \\ \omega Ax &= \omega b \\ (\omega D + \omega L + \omega U)x &= \omega b \\ (D + \omega L)x &= \omega b - \omega Ux + (1-\omega)Dx \\ (D + \omega L)x_{k+1} &= \omega b - \omega Ux_k + (1-\omega)Dx_k \\ Dx_{k+1} &= \omega b + (1-\omega)Dx_k - \omega Ux_k - \omega Lx_{k+1} \\ x_{k+1} &= (1-\omega)x_k + D^{-1}(\omega b - \omega Ux_k - \omega Lx_{k+1}) \end{aligned}$$

$$\begin{aligned} x_0 &= \text{Initial Vector} \\ x_{k+1} &= (1-\omega)x_k + D^{-1}(\omega b - \omega Ux_k - \omega Lx_{k+1}) \\ x_{k+1,i} &= (1-\omega)x_{k,i} + \frac{\omega}{a_{ii}}(b_i - \sum_{j>i} a_{ij}x_{k,j} - \sum_{j<i} a_{ij}x_{k+1,j}) \textbf{ for } 1 \le i \le n \end{aligned}$$

共轭梯度法

- **计算过程**（如果算法流程给出的话直接套公式就行）
  - 符号：
    - d_k：第k个互共轭向量，代表前进方向
    - α_k：d_k和x^*的系数，确保$(d_k, r_{k+1}) = 0$，代表步长
    - x_k：在第k步时取得的近似解
      - x^*在$\{d_1, \ldots, d_{k-1}\}$的投影，即$\sum_{i=1}^{k-1} \alpha_i d_i$
    - r_k：在第k步时，x_k的残差，即$b - Ax_k$
      - r_k满足$(r_i, r_k) = 0$，$0 \le i \le k$
    - β_k：保证$(d_k, d_{k+1})_A = 0$的系数
  - 算法描述：

$$
\begin{aligned}
& x_0 = \text{Initial Guess} \\
& d_0 = r_0 = b - Ax_0 \\
& \textbf{for } k = 0, 1, 2, \ldots, n-1 \textbf{ do} \\
& \quad \textbf{if } r_k \textit{ is sufficiently small } \textbf{then} \\
& \quad \quad \textbf{return } x_k \\
& \quad \alpha_k = \frac{r_k^\top r_k}{d_k^\top A d_k} \\
& \quad x_{k+1} = x_k + \alpha_k d_k \\
& \quad r_{k+1} = r_k - \alpha_k A d_k \\
& \quad \beta_k = \frac{r_{k+1}^\top r_{k+1}}{r_k^\top r_k} \\
& \quad d_{k+1} = r_{k+1} + \beta_k d_k
\end{aligned}
$$

- **主要定理（三个）**，给出α_k，d_k等算法中会用到的五个符号
  - 前置条件：Let $b \neq 0$, $x_0 = 0$, and $r_k \neq 0$ for $k < n$. Then for each $1 \leq k \leq n$,

1. **The following three subspace of R^n are equal**:

$$(x_1, \ldots, x_k) = (r_0, \ldots, r_{k-1}) = (d_0, \ldots, d_{k-1})$$

- **Proof of 1st item**:
  - **Base case** $(k = 1)$: $(x_1) = (r_0) = (d_0)$ since $x_1 = x_0 + \alpha d_0$, $d_0 = r_0 = b - Ax_0$
  - **Inductive step** $(k > 1)$: Suppose that the $k-1$ case hold, prove $(x_1, \ldots, x_k) = (d_0, \ldots, d_{k-1})$:
    - $(x_1, \ldots, x_k) \subseteq (d_0, \ldots, d_{k-1})$ since $x_k = \sum_{i=0}^{k-1} \alpha_i d_i$;
    - $(x_1, \ldots, x_k) \supseteq (d_0, \ldots, d_{k-1})$ since $x_k = x_{k-1} + \alpha_{k-1} d_{k-1} \Rightarrow d_{k-1} = \dfrac{1}{\alpha_{k-1}} x_k - \dfrac{1}{\alpha_{k-1}} x_{k-1}$
  - Prove $(r_0, \ldots, r_{k-1}) = (d_0, \ldots, d_{k-1})$:
    - $(r_0, \ldots, r_{k-1}) \subseteq (d_0, \ldots, d_{k-1})$ since $d_{k-1} = r_{k-1} + \beta_{k-2} d_{k-2} \Rightarrow r_{k-1} = d_{k-1} - \beta_{k-2} d_{k-2}$
    - $(r_0, \ldots, r_{k-1}) \supseteq (d_0, \ldots, d_{k-1})$ since $d_{k-2} = \sum_{i=0}^{k-2} \gamma_i r_i \Rightarrow d_{k-1} = r_{k-1} + \beta_{k-2} d_{k-2} = r_{k-1} + \sum_{i=0}^{k-2} (\beta_{k-2} \gamma_i) r_i$

2. **Distinct residuals are pairwise orthogonal**:

$$r_k^\top r_j = 0 \text{ for } j < k$$

- **Proof of 2nd item**:
  - 前置引理**Lemma 1**: $(r_j, d_k)_A = 0$ for $0 \leq j < k$ or $0 \leq k < j+1$
    - **Proof of lemma 1**:
      Here only prove the case that $0 \leq j < k$:
      - If $j = 0$, then $d_k^\top A r_0 = d_k^\top A d_0 = 0$
      - Otherwise, $d_k^\top A r_j = d_k^\top A (d_j - \beta_{j-1} d_{j-1}) = d_k^\top A d_j - \beta_{j-1} d_k^\top A d_{j-1} = 0$
  - 前置引理**Lemma 2**: $d_k^\top A d_k = r_k^\top A d_k$
    - **Proof of lemma 2**:
      $d_k^\top A d_k = (r_k^\top + \beta_{k-1} d_{k-1}^\top) A d_k = r_k^\top A d_k + \beta_{k-1} d_{k-1}^\top A d_k = r_k^\top A d_k$
  - **Base case** $(k = 1)$:
    - $r_0^\top r_1 = r_0^\top (r_0 - \alpha_0 A d_0) = r_0^\top r_0 - \alpha_0 r_0^\top A d_0 = r_0^\top r_0 - \dfrac{r_0^\top r_0}{d_0^\top A d_0} d_0^\top A d_0 = 0$
  - **Inductive step** $(k > 1)$: Suppose that the $k-1$ case hold:
    - $r_j^\top r_k = r_j^\top r_{k-1} - \alpha_{k-1} r_j^\top A d_{k-1}$
    - If $j < k-1$, then $r_j^\top r_k = 0$
    - If $j = k-1$, then $\alpha_{k-1} = \dfrac{r_{k-1}^\top r_{k-1}}{d_{k-1}^\top A d_{k-1}}$
    - $r_{k-1}^\top r_k = r_{k-1}^\top (r_{k-1} - \alpha_{k-1} A d_{k-1}) = r_{k-1}^\top r_{k-1} - \alpha_{k-1} r_{k-1}^\top A d_{k-1} = r_{k-1}^\top r_{k-1} - \dfrac{r_{k-1}^\top r_{k-1}}{d_{k-1}^\top A d_{k-1}} d_{k-1}^\top A d_{k-1} = 0$

3. **Distinct vectors of a subspace span are pairwise A-conjugate**:

$$d_k^\top A d_j = 0 \text{ for } j < k$$

- **Proof of 3rd item**:
  - **Base case** $(k = 1)$:
    - $\beta_0 = -\dfrac{r_1^\top A d_0}{d_0^\top A d_0}$
    - $d_0^\top A d_1 = d_0^\top A (r_1 + \beta_0 d_0) = d_0^\top A r_1 + \beta_0 d_0^\top A d_0 = d_0^\top A r_1 - \dfrac{r_1^\top A d_0}{d_0^\top A d_0} d_0^\top A d_0 = 0$

- **Inductive step** $(k>1)$: Suppose that the $k-1$ case hold:
  - $d_j^\top Ad_k = d_j^\top Ar_k + \beta_{k-1}d_j^\top Ad_{k-1}$
  - If $j<k-1$, then $Ad_j = \dfrac{r_j - r_{j+1}}{\alpha_j}$ is orthogonal to r_k, i.e., $d_j^\top Ar_k = 0$
  - So $d_j^\top Ad_k = 0$
  - If $j = k-1$, then $\beta_{k-1} = -\dfrac{r_k^\top Ad_{k-1}}{d_{k-1}^\top Ad_{k-1}}$
  - So $d_{k-1}^\top Ad_k = d_{k-1}^\top A(r_k + \beta_{k-1}d_{k-1}) = d_{k-1}^\top Ar_k + \beta_{k-1}d_{k-1}^\top Ad_{k-1} = d_{k-1}^\top Ar_k - \dfrac{r_k^\top Ad_{k-1}}{d_{k-1}^\top Ad_{k-1}}d_{k-1}^\top Ad_{k-1} = 0$

- 具有预处理的共轭梯度法
  - 前提条件：A is ill-conditioned: the condition number of A is very large.
  - 定义：
    设$M=M_1M_2$是非奇异的且$Ax=b$是一个线性方程组，我们令$\tilde{A}\tilde{x}=\tilde{b}$是一个线性方程组，此时矩阵$M$就是一个预调节器（**preconditioner**），其中：
    - $\tilde{A} = M_1^{-1}AM_2^{-1}$
    - $\tilde{x} = M_2x$
    - $\tilde{b} = M_1^{-1}b$
  - 三种预调节器，令$A = L + D + L^\top$：
    - Jacobi: $M = D$
    - Gauss-Seidel: $M = (D+L)D^{-1}(D+L)^\top$
    - SOR: $M = (D+\omega L)D^{-1}(D+\omega L)^\top$ where $0 \le \omega \le 2$
  - 如果M是一个对称正定矩阵，那么其存在一个唯一的对称正定矩阵C使得$M=C^2$，因此利用预处理的共轭梯度法会将线性方程组处理为：
    - $\tilde{A} = C^{-1}AC^{-1}$
    - $\tilde{x} = Cx$
    - $\tilde{b} = C^{-1}b$
  - 算法描述：
    - 引入了新符号z_k，其为一个辅助向量

$$
\begin{aligned}
&x_0 = \text{Initial Guess}\\
&r_0 = b - Ax_0\\
&d_0 = z_0 = M^{-1}r_0\\
&\textbf{for } k = 0,1,2,\ldots,n-1 \textbf{ do}\\
&\quad \textbf{if } r_k \textit{ is sufficiently small } \textbf{then}\\
&\quad\quad \textbf{return } x_k\\
&\quad \tilde{\alpha}_k = \frac{r_k^\top z_k}{d_k^\top Ad_k}\\
&\quad x_{k+1} = x_k + \tilde{\alpha}_k d_k\\
&\quad r_{k+1} = r_k - \tilde{\alpha}_k Ad_k\\
&\quad z_{k+1} = M^{-1}r_{k+1}\\
&\quad \tilde{\beta}_k = \frac{r_{k+1}^\top z_{k+1}}{r_k^\top z_k}\\
&\quad d_{k+1} = z_{k+1} + \tilde{\beta}_k d_k
\end{aligned}
$$