Solución de sistemas lineales

PhD. Alejandro Paredes

Facultad de Ciencias

24 de abril de 2023

Resumen

1. Métodos directos e iterativos
2. Métodos iterativos
   - Método de Jacobi
   - Método Gauss-Seidel
   - SOR
3. Métodos de optimización
   - Método de máximo decenso
   - Método de gradiente conjugado
4. Sistemas mal condicionados.

Métodos directos e iterativos

Según las características del problema

- Tipo de matriz (triangular, tridiagonal, llena de ceros, sin ceros, etc)
- Tamaño del problema (dimensión de la matriz).
- Condicionamiento de la matriz.

Existen dos métodos (caminos) para encontrar la solución

- Métodos directos(Sustitución directa e inversa, eliminación de Gauss con y sin pivote).
- Métodos iterativos(Jacobi, Gaus-Seidel,SOR, método del gradiente).

Nota: En ningún caso se calcula A^{-1}

Métodos iterativos $\mathcal{O}(n^2)$

- Adaptados para sistemas de gran dimensión.
- Calcula soluciones aproximadas.

Sistema lineal a resolver:
Solución : $[-1 \quad 3 \quad 2]$

$$\begin{array}{ccc} \mathbf{A} & \mathbf{x} & = & \mathbf{b} \\ \begin{bmatrix} 4 & 2 & 3 \\ 3 & -5 & 2 \\ -2 & 3 & 8 \end{bmatrix} & \begin{bmatrix} x \\ y \\ z \end{bmatrix} & = & \begin{bmatrix} 8 \\ -14 \\ 27 \end{bmatrix} \end{array}$$

$$\begin{array}{ccccccc} \mathbf{A} & = & \mathbf{D} & + & \mathbf{L} & + & \mathbf{U} \\ \begin{bmatrix} 4 & 2 & 3 \\ 3 & -5 & 2 \\ -2 & 3 & 8 \end{bmatrix} & = & \begin{bmatrix} 4 & 0 & 0 \\ 0 & -5 & 0 \\ 0 & 0 & 8 \end{bmatrix} & + & \begin{bmatrix} 0 & 0 & 0 \\ 3 & 0 & 0 \\ -2 & 3 & 0 \end{bmatrix} & + & \begin{bmatrix} 0 & 2 & 3 \\ 0 & 0 & 2 \\ 0 & 0 & 0 \end{bmatrix} \end{array}$$

Método de Jacobi

Si no hay ceros en la diagonal principal, entonces:

$$\begin{aligned} x &= (8 - 2y - 3z)4 \\ y &= (-14 - 3x - 2z)/(-5) \\ z &= (27 + 2x - 3y)/8 \end{aligned}$$

Los nuevos valores de x,y,z se obtienen a partir de los antiguos

$$\begin{aligned} x^{k+1} &= (8 - 2y^k - 3z^k)4 \\ y^{k+1} &= (-14 - 3x^k - 2z^k)/(-5) \\ z^{k+1} &= (27 + 2x^k - 3y^k)/8 \end{aligned}$$

con $(x^0, y^0, z^0) = (0, 0, 0)$

Método de Jacobi

$$\mathbf{D}\mathbf{x}^{k+1} = \mathbf{b} - (\mathbf{L} + \mathbf{U})\mathbf{x}^{k}$$

Forma compacta:

$$\mathbf{x}^{k+1} = \mathbf{D}^{-1}\left[\mathbf{b} - (\mathbf{L} + \mathbf{U})\mathbf{x}^{k}\right]$$

Formula de recurrencia:

$$x_i^{(k+1)} = \frac{1}{a_{ii}}\left(b_i - \sum_{i \neq j} a_{ij} x_j^{(k)}\right)$$

Método Gauss-Seidel

A partir del método de Jacobi

$$
\begin{aligned}
x^{k+1} &= (8 - 2y^k - 3z^k)4 \\
y^{k+1} &= (-14 - 3x^k - 2z^k)/(-5) \\
z^{k+1} &= (27 + 2x^k - 3y^k)/8
\end{aligned}
$$

se actualizan los valores que ya se conocen

$$
\begin{aligned}
x^{k+1} &= (8 - 2y^k - 3z^k)4 \\
y^{k+1} &= (-14 - 3{\color{red}x^{k+1}} - 2z^k)/(-5) \\
z^{k+1} &= (27 + 2{\color{red}x^{k+1}} - 3{\color{red}y^{k+1}})/8
\end{aligned}
$$

es más eficiente.

Método Gauss-Seidel-Forma compacta

El método de Gauss-Seidel se puede expresar de la siguiente manera:

$$(\mathbf{D}+\mathbf{L})\mathbf{x}^{k+1} = \mathbf{b} - \mathbf{U}\mathbf{x}^{k}$$

Forma compacta:

$$\mathbf{x}^{k+1} = (\mathbf{D}+\mathbf{L})^{-1}\left[\mathbf{b} - \mathbf{U}\mathbf{x}^{k}\right]$$

Formula de recurrencia:

$$x_i^{(k+1)} = \frac{1}{a_{ii}}\left(b_i - \sum_{j=i+1}^{n} a_{ij}x_j^{(k)} - \sum_{j=1}^{i-1} a_{ij}x_j^{(k+1)}\right)$$

Sucesive Over Relaxation (SOR)

Descomposición de **A** :

$$\left[\mathbf{D} + \omega\mathbf{L} + (1-\omega)\mathbf{L} + \mathbf{U}\right]\mathbf{x}^{k+1} = \mathbf{b}$$

Forma compacta:

$$\left[\mathbf{D} + \omega\mathbf{L}\right]\mathbf{x}^{k+1} = \mathbf{b} - \left[(1-\omega)\mathbf{L} + \mathbf{U}\right]\mathbf{x}^{k}$$

Formula de recurrencia

$$x_i^{(k+1)} = \frac{1}{a_{ii}} \left(b_i - (1-\omega)\sum_{j=1}^{i-1} a_{ij}x_j^{(k)} - \sum_{j=i+1}^{n} a_{ij}x_j^{(k)} - \omega\sum_{j=1}^{i-1} a_{ij}x_j^{(k+1)} \right)$$

Métodos iterativos estacionarios

A partir del sistema original $\mathbf{x} = G\mathbf{x} + c$, se propone el método iterativo de la forma

$$\mathbf{x}_{k+1} = G\mathbf{x}_k + c$$

por lo tanto tenemos

- Jacobi: $G = -D^{-1}(L + U)$
- Gauss-Seidel: $G = -(D + L)^{-1}U$
- SOR $G = -\left[\mathbf{D} + \omega\mathbf{L}\right]^{-1}\left[(1-\omega)\mathbf{L} + \mathbf{U}\right]$

Error en la $n - esima$ aproximación es $\mathbf{e}^n = \mathbf{x} - \mathbf{x}^n$

$$\begin{aligned} \mathbf{e}^{n+1} &= G\mathbf{e}^n \\ \mathbf{e}^n &= G\mathbf{e}^{n-1} = G^2\mathbf{e}^{n-2} = \cdots = G^n\mathbf{e}^0 \end{aligned}$$

La secuencia de soluciones $\{\mathbf{x}_1, \mathbf{x}_2, \dots\}$ converge cuando

$$\lim_{n\to\infty} \mathbf{e}^n = \lim_{n\to\infty} G^n\mathbf{e}^0 = \mathbf{0} \quad \Rightarrow \quad \boxed{\lim_{n\to\infty} G^n = \mathbf{0}}$$

Métodos iterativos estacionarios

Si $dim(G)$ es m, suponemos que G tiene m autovectores independientes $\mathbf{v}_s (s = 1, \ldots, m)$ y λ_s autovectores. Los $\mathbf{v}_s$ forman una base y

$$\mathbf{e}^0 = \sum_{s=1,m} d_s \mathbf{v}_s$$

Además

$$\mathbf{e}^1 = \sum_{s=1,m} d_s G \mathbf{v}_s = \sum_{s=1,m} d_s \lambda_s \mathbf{v}_s \quad \Rightarrow \quad \boxed{\mathbf{e}^n = \sum_{s=1,m} d_s \lambda_s^n \mathbf{v}_s}$$

para un $\mathbf{e}^0$ arbitrario, $\mathbf{e}^n$ tenderá a un vector nulo si y sólo si $|\lambda_s < 1| \; \forall s$.
El proceso convergerá para un $\mathbf{x}^0$ arbitrario si el radio espectral $\rho(G) < 1$

Método de optimización

Considerereremos

$$\Phi(x) = \frac{1}{2}x^T Ax - x^T b$$

donde $b, x \in R^n$ y $A \in R^{nn}$ es simétrica positiva definida (SPD)
Nota: SPD equivale a decir $x^T Ax > 0 \quad \forall x$ o los autovalores de A son positivos.
Sea x^* el valor que minimiza $\Phi(x)$

$$\nabla\Phi(x^*) = Ax^* - b = 0 \quad \Rightarrow \quad Ax^* = b$$

Resolver el sistema lineal se ha convertido en un problema de optimización

Método de máximo decenso

Planteamos que la solución se busca de manera iterativa:

$$x_{k+1} = x_k + \alpha_k p_k; \quad p_k \text{ : vector}, \quad \alpha_k \text{ : escalar}$$

se escoge una solución incial x_0 tal que $\Phi(x_{k+1}) < \Phi(x_k)$.
Sabemos que:

- $\nabla\Phi(x)$: vector con direccón de máximo incremento de la función $\Phi(x)$.
- $-\nabla\Phi(x)$: localmente la dirección de máximo decenso $p_k = -\nabla\Phi(x)$.

$$\begin{aligned} x_{k+1} &= x_k - \alpha_k \nabla\Phi(x_k), \quad \nabla\Phi(x_k) = Ax_k - b = r_k \\ \Phi(x_{k+1}) &= \frac{1}{2}\left(x_k - \alpha_k r_k\right)^T A \left(x_k - \alpha_k r_k\right) - \left(x_k - \alpha_k r_k\right)^T b \end{aligned}$$

Método de máximo decenso

Para determinar α_k, minimizamos $\Phi(x_{k+1})$ con respecto de α_k

$$
\begin{aligned}
\Phi(x_{k+1}) &= \frac{1}{2}\left(x_k - \alpha_k r_k\right)^T A \left(x_k - \alpha_k r_k\right) - \left(x_k - \alpha_k r_k\right)^T b \\
\frac{d}{d\alpha_k}\Phi(x_{k+1}) &= r_k^T A x_k + \alpha_k r_k^T A r_k - r_k^T b = 0 \\
\alpha_k &= \frac{\nabla\Phi(x_k)^T \nabla\Phi(x_k)}{\nabla\Phi(x_k)^T A \nabla\Phi(x_k)} = \frac{r_k^T r_k}{r_k^T A r_k} = \frac{\langle r_k, r_k\rangle}{\langle r_k, A r_k\rangle}
\end{aligned}
$$

Método de gradiente conjugado

Conjunto de vectores conjugados $\{p_0, p_1, p_2, \ldots, p_{n-1}\}$ con respecto a la matriz A

$$p_i^T A p_j = 0 \quad i \neq j$$

forman una base de R^n.
Definimos :

$$\begin{aligned} x_{k+1} &= x_k + \alpha_k p_k \\ p_{k+1} &= -r_{k+1} + \beta_k p_k, \quad \text{donde } r_k = Ax_k - b \end{aligned}$$

- Se dede determinar α_k y β_k.
- Se puede demostrar que $r_k^T r_j = 0 \quad \forall k \neq j$

Método de gradiente conjugado

Para determinar α_k se minimiza la función $\Phi(x_{k+1})$

$$\begin{aligned}\Phi(x_k+1) &= \frac{1}{2}\left(x_k+\alpha_k p_k\right)^T\left(x_k+\alpha_k p_k\right)-\left(x_k+\alpha_k p_k\right)^T b \\ \frac{d}{d\alpha_k}\Phi(x_{k+1}) &= p_k^T r_k+\alpha_k p_k^T A p_k = 0 \Rightarrow \boxed{\alpha_k = -\frac{p_k^T r_k}{p_k^T A p_k}}\end{aligned}$$

usando $p_k = -r_k + \beta_{k-1} p_{k-1}$, tenemos

$$\alpha_k = \frac{r_k^T r_k}{p_k^T A p_k} - \beta_{k-1}\frac{p_{k-1}^T r_k}{p_k^T A p_k}$$

degradando sucesivamente $p_{k-1} \approx r_{k-1}$ y usando la ortogonalidad de los r_k, tenemos

$$\boxed{\alpha_k = \frac{r_k^T r_k}{p_k^T A p_k}}$$

Método de gradiente conjugado

Para determinar β_k, utilizamos

$$\begin{aligned} p_k^T A(p_{k+1} &= -r_{k+1} + \beta_k p_k) \quad \text{sabiendo que } p_k^T A p_{k+1} = 0 \\ \beta_k &= \frac{p_k^T A r_{k+1}}{p_k A p_k} \end{aligned}$$

utilizando las definiciones de p_k y usando la ortogonalidad de los r_k se puede mostrar que

$$\boxed{\beta_k = \frac{r_{k+1}^T r_{k+1}}{r_k^T r_k}}$$

Método de gradiente conjugado

For $k = 0, 1, 2, ..$ until convergence

$$\alpha_k = \frac{{r_k}^T r_k}{{p_k}^T A p_k}$$
$$x_{k+1} = x_k + \alpha_k p_k$$
$$r_{k+1} = r_k + \alpha_k A p_k$$
$$\beta_k = \frac{{r_{k+1}}^T r_{k+1}}{{r_k}^T r_k}$$
$$p_{k+1} = -r_{k+1} + \beta_k p_k$$

End

Método de gradiente conjugado

Para $k = 0$ tenemos

$$
\begin{aligned}
\alpha_0 &= \frac{r_0^T r_0}{p_0^T A p_0} \\
x_1 &= x_0 + \alpha_0 p_0 \\
r_1 &= r_0 + \alpha_0 A p_0 \\
\beta_0 &= \frac{r_1^T r_1}{r_0^T r_0} \\
p_1 &= -r_1 + \beta_0 p_0
\end{aligned}
$$

- Necesitamos $r_0 = Ax_0 - b$ y p_0.
- x_0 y p_0 son arbitrarios.
- Tomamos x_0 cualquiera (inclusive cero) y $p_0 = -r_0$.

Estimación del error

Error en la iteración i:

$$\delta_i = |\mathbf{X}^i - \mathbf{X}^{i-1}|, \quad i > 1$$

Criterio de convergencia:

$$\text{Si } \delta_i < \epsilon \quad \Rightarrow \quad \text{detener método}$$

Sistemas mal condicionados

Ref: [Ascher, First Course in Numerical Methods]
Si tenemos un sistema lineal

$$A\mathbf{x} = \mathbf{b},$$

donde $\mathbf{x}$ es la solución exacta y $\hat{x}$ es la solución aproximada.
Sería deseable calcular, para alguna norma, el error

- absoluto : $||\mathbf{x} - \hat{x}||$
- relativo $||\mathbf{x} - \hat{x}||/||\mathbf{x}||$

Ningún error se puede calcular, mejor es analizar el residuo

$$\hat{r} = \mathbf{b} - A\hat{x}$$

Sistemas mal condicionados

Estimación del error relativo

$$\begin{aligned} \hat{r} &= \mathbf{b} - A\hat{x} = A\mathbf{x} - A\hat{x} = A(\mathbf{x} - \hat{x}) \\ \mathbf{x} - \hat{x} &= A^{-1}\hat{r} \end{aligned}$$

para una norma $||, ||$ y la correspondiente norma matricial inducida

$$||\mathbf{x} - \hat{x}|| = ||A^{-1}\hat{r}|| \leq ||A^{-1}|| \quad ||\hat{r}|| \quad \text{error absoluto acotado}$$

por otro lado tenemos $||\mathbf{b}|| \leq ||A|| \quad ||\mathbf{x}||$

$$\frac{||\mathbf{x} - \hat{x}||}{||\mathbf{x}||} \leq \kappa(A) \frac{||\hat{r}||}{||\mathbf{b}||}$$

donde $\kappa(A) = ||A|| \quad ||A^{-1}||$ es el número de condicionamiento.

Sistemas mal condicionados

Observaciones

- $\kappa(A)$ no depende de la norma utilizada.
- El error relativo esta acotado por el valor de $\kappa(A)$ multiplicado por el residuo.
- $\kappa(A) \approx 1$ se dice que el sistema esta bien condicionado.
- $\kappa(A) \gg 1$ se dice que el sistema esta mal condicionado y ningún método encontrará la solución.

Sistemas mal condicionados

Ejemplo

$$A = \begin{bmatrix} 1{,}2969 & 0{,}8648 \\ 0{,}2161 & 0{,}1441 \end{bmatrix} \quad ; \quad b = \begin{bmatrix} 0{,}8642 \\ 0{,}1440 \end{bmatrix}$$

$$\hat{x} = \begin{bmatrix} 0{,}9911 \\ 0{,}4870 \end{bmatrix} \quad , \qquad \hat{r} = \mathbf{b} - A\hat{x} = \begin{bmatrix} -10^{-8} \\ 10^{-8} \end{bmatrix}$$

considerando la norma infinita $||x||_\infty = max_{1\leq i\leq n}|x_i|$ y sabiendo que

$$\mathbf{x} = \begin{bmatrix} 2 \\ -2 \end{bmatrix}, \text{ entonces } ||\mathbf{x} - \hat{x}||_\infty = 1{,}513$$

El error absoluto es del mismo orden que la solución $\mathbf{x}$

Sistemas mal condicionados

$$A^{-1} = 10^8 \left[\begin{array}{cc} 0{,}1441 & -0{,}8648 \\ -0{,}2161 & 1{,}2969 \end{array} \right] \quad ; \quad ||A||_\infty = max_{1 \leq i \leq m} \sum_{j=1}^{n} |a_{ij}|$$

$$||A^{-1}||_\infty = 1{,}513 \times 10^8 \quad , \quad \kappa_\infty(A) = 3{,}27 \times 10^8$$

tenemos que $\kappa_\infty(A) \gg 1$, entonces la matriz esta mal condicionada.