MAT-468: Sesión 3,
Solución de sistemas lineales II

Felipe Osorio

http://fosorios.mat.utfsm.cl

Departamento de Matemática, UTFSM

EX UMBRA IN SOLEM

Métodos iterativos

Considere resolver el sistema lineal

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

donde $\boldsymbol{A} \in \mathbb{R}^{n \times n}$ y asumiremos que el sistema admite una solución exacta $\widehat{\boldsymbol{x}}$.

A continuación se describe:

- El método Jacobi.
- Método Gauss-Seidel.
- Método de gradientes conjugados.

Para uniformizar la presentación de todos los métodos, sea $\boldsymbol{D}$, $\boldsymbol{L}$ y $\boldsymbol{U}$ las partes diagonal, triangular inferior y triangular superior de la matriz $\boldsymbol{A}$, respectivamente[1].

Es decir $\boldsymbol{D} = \mathrm{diag}(a_{11}, a_{22}, \ldots, a_{nn})$, mientras que

$$\boldsymbol{L} = \begin{pmatrix} 0 & 0 & 0 & \ldots & 0 \\ a_{21} & 0 & 0 & \ldots & 0 \\ a_{31} & a_{32} & 0 & \ldots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \ldots & a_{n,n-1} & 0 \end{pmatrix}, \qquad \boldsymbol{U} = \begin{pmatrix} 0 & a_{12} & a_{13} & \ldots & a_{1n} \\ 0 & 0 & a_{23} & \ldots & a_{2n} \\ \vdots & \vdots & \ddots & & \vdots \\ 0 & 0 & \ldots & 0 & a_{n-1,n} \\ 0 & 0 & \ldots & 0 & 0 \end{pmatrix}.$$

[1] Evidentemente $\boldsymbol{A} = \boldsymbol{D} + \boldsymbol{L} + \boldsymbol{U}$.

Un método iterativo para la solución de sistemas lineales construye una secuencia $\{\boldsymbol{x}^{(k)}\}$, para $k = 0, 1, \ldots,$ que bajo ciertas condiciones converge a $\widehat{\boldsymbol{x}}$.

$\boldsymbol{x}^{(0)}$ representa una estimación inicial y la secuencia es construída como una regla del tipo:

$$\boldsymbol{x}^{(k+1)} = \boldsymbol{B}_k \boldsymbol{x}^{(k)} + \boldsymbol{C}_k \boldsymbol{b}, \quad k = 0, 1, \ldots, \tag{1}$$

donde $\boldsymbol{B}_k, \boldsymbol{C}_k \in \mathbb{R}^{n \times n}$.

Las diferentes elecciones de $\boldsymbol{B}_k$ y $\boldsymbol{C}_k$ definen los distintos algoritmos.

La condición básica sobre $\boldsymbol{B}_k$ y $\boldsymbol{C}_k$ que asegura la convergencia del algoritmo es la siguiente:

$$\boldsymbol{B}_k + \boldsymbol{C}_k \boldsymbol{A} = \boldsymbol{I}.$$

En efecto, se debe tener que

$$\widehat{\boldsymbol{x}} = \boldsymbol{B}_k \widehat{\boldsymbol{x}} + \boldsymbol{C}_k \boldsymbol{b},$$

y dado que $\widehat{\boldsymbol{x}}$ es solución del sistema de ecuaciones $\boldsymbol{A}\boldsymbol{x} = \boldsymbol{b}$, tenemos

$$\begin{aligned}\widehat{\boldsymbol{x}} &= \boldsymbol{B}_k \widehat{\boldsymbol{x}} + \boldsymbol{C}_k \boldsymbol{A} \widehat{\boldsymbol{x}} \\ &= (\boldsymbol{B}_k + \boldsymbol{C}_k \boldsymbol{A}) \widehat{\boldsymbol{x}},\end{aligned}$$

desde donde sigue el resultado.

Resultado 1

Una secuencia $\{\boldsymbol{x}^{(k)}\}$ dada por (1) converge a la solución del sistema $\boldsymbol{A}\boldsymbol{x} = \boldsymbol{b}$ para cualquier $\boldsymbol{x}^{(0)}$, sólo si

$$\lim_{k\to\infty} \boldsymbol{B}_k \boldsymbol{B}_{k-1} \cdots \boldsymbol{B}_0 = \boldsymbol{0}$$

En la práctica se utiliza métodos iterativos estacionarios, es decir, tal que $\boldsymbol{B}_k = \boldsymbol{B}$ y $\boldsymbol{C}_k = \boldsymbol{C}$ son constantes. De este modo, tenemos la siguiente regla:

$$\boldsymbol{x}^{(k+1)} = \boldsymbol{B}\boldsymbol{x}^{(k)} + \boldsymbol{C}\boldsymbol{b}, \quad k = 0, 1, \dots, \tag{2}$$

y la condición de convergencia en el resultado anterior adopta una forma más simple

Resultado 2

Una secuencia $\{\boldsymbol{x}^{(k)}\}$ dada por (2) converge a la solución del sistema $\boldsymbol{A}\boldsymbol{x} = \boldsymbol{b}$ para cualquier $\boldsymbol{x}^{(0)}$, si y sólo si el radio espectral $\rho(\boldsymbol{B}) < 1$, donde $\rho(\boldsymbol{B}) = \max\{\lambda_1, \dots, \lambda_n\}$ con $\lambda_1, \dots, \lambda_n$ los valores propios de $\boldsymbol{B}$[2].

[2] Esta condición es equivalente a $\|\boldsymbol{B}\| < 1$ para cualquier norma matricial.

Criterio de parada

Se suele declarar convergencia cuando algún criterio preestablecido es satisfecho.

Específicamente, para un nivel de tolerancia τ, el proceso iterativo se detiene si

$$\|\boldsymbol{x}^{(k+1)} - \boldsymbol{x}^{(k)}\| \leq \tau, \qquad \|\boldsymbol{r}^{(k+1)} - \boldsymbol{r}^{(k)}\| \leq \tau, \quad \text{o} \quad \|\boldsymbol{r}^{(k+1)}\| \leq \tau,$$

donde $\boldsymbol{r}^{(k)} = \boldsymbol{A}\boldsymbol{x}^{(k)} - \boldsymbol{b}$ es el vector residual.

Además se suele incluir un número máximo aceptable de iteraciones.

Suponga que $\boldsymbol{A}$ tiene elementos diagonales no cero. Entonces el sistema $\boldsymbol{Ax} = \boldsymbol{b}$ puede ser escrito como

$$\boldsymbol{Dx} + (\boldsymbol{L} + \boldsymbol{U})\boldsymbol{x} = \boldsymbol{b},$$

y de este modo, tenemos que

$$\boldsymbol{x} = \boldsymbol{D}^{-1}[-(\boldsymbol{L} + \boldsymbol{U})\boldsymbol{x} + \boldsymbol{b}]$$

Esto motiva el siguiente algoritmo:

$$\boldsymbol{x}^{(k+1)} = -\boldsymbol{D}^{-1}(\boldsymbol{L} + \boldsymbol{U})\boldsymbol{x}^{(k)} + \boldsymbol{D}^{-1}\boldsymbol{b}, \quad k = 0, 1, \dots, \tag{3}$$

Observación:

La condición $\rho(\boldsymbol{D}^{-1}(\boldsymbol{L} + \boldsymbol{U})) < 1$ es satisfecha para una amplia clase de matrices (por ejemplo, diagonal dominantes, simétricas y definidas positivas).

Análogamente al método Jacobi, podemos reescribir el sistema $\boldsymbol{A}\boldsymbol{x} = \boldsymbol{b}$ como

$$(\boldsymbol{L} + \boldsymbol{D})\boldsymbol{x} + \boldsymbol{U}\boldsymbol{x} = \boldsymbol{b},$$

es decir

$$\boldsymbol{x} = (\boldsymbol{L} + \boldsymbol{D})^{-1}[-\boldsymbol{U}\boldsymbol{x} + \boldsymbol{b}]$$

Esto lleva al esquema iterativo:

$$\boldsymbol{x}^{(k+1)} = -(\boldsymbol{L} + \boldsymbol{D})^{-1}\boldsymbol{U}\boldsymbol{x}^{(k)} + (\boldsymbol{L} + \boldsymbol{D})^{-1}\boldsymbol{b}, \quad k = 0, 1, \ldots, \tag{4}$$

Observación:

La condición de convergencia $\rho((\boldsymbol{L} + \boldsymbol{D})^{-1}\boldsymbol{U}) < 1$ es satisfecha para matrices diagonal dominantes y definidas positivas.

EX UMBRA IN SOLEM

El método Gauss-Seidel puede ser inaceptablemente lento. El método de sobrerelajación sucesiva (SOR) está basado en la siguiente identidad

$$(\boldsymbol{D} + \omega\boldsymbol{L})\boldsymbol{x} = -[\omega\boldsymbol{U} + (\omega - 1)\boldsymbol{D}]\boldsymbol{x} + \omega\boldsymbol{b},$$

donde $\omega > 0$ es llamado factor de relajación. Se puede mostrar que SOR converge solamente para valores de $\omega \in (0, 2)$ (ver Kahan, 1958).

De este modo, obtenemos:

$$\begin{aligned} \boldsymbol{x}^{(k+1)} &= (\boldsymbol{D} + \omega\boldsymbol{L})^{-1}[(\omega - 1)\boldsymbol{D} - \omega\boldsymbol{U}]\boldsymbol{x}^{(k)} + \omega(\boldsymbol{D} + \omega\boldsymbol{L})^{-1}\boldsymbol{b}, \\ &= \boldsymbol{M}_{\omega}\boldsymbol{x}^{(k)} + \boldsymbol{d}_{\omega}, \qquad k = 0, 1, \dots, \end{aligned} \tag{5}$$

Observación:

Una buena elección del parámetro ω[3] puede acelerar la convergencia del método[4] (Young, 1954; Hadjidimos, 2000).

[3] $\omega_{\text{opt}} = 2/(1 + \rho(\boldsymbol{B}))$, con $\boldsymbol{B} = -(\boldsymbol{L} + \boldsymbol{D})^{-1}\boldsymbol{U}$.

[4] Un radio espectral menor, indica convergencia más rápida

Método de gradientes conjugados

Resolver el sistema lineal $\boldsymbol{A}\boldsymbol{x} = \boldsymbol{b}$ es equivalente a determinar el mínimo de la función:

$$\phi(\boldsymbol{x}) = \frac{1}{2}\boldsymbol{x}^\top \boldsymbol{A}\boldsymbol{x} - \boldsymbol{x}^\top \boldsymbol{b}.$$

En efecto, por hacer la derivada de ϕ igual a $\boldsymbol{0}$, vemos que el punto estacionario de ϕ ocurre en el punto $\boldsymbol{x}$ donde $\boldsymbol{A}\boldsymbol{x} = \boldsymbol{b}$.

- Este procedimiento asume que la matriz $\boldsymbol{A}$ es simétrica y definida positiva.
- En este caso, el (único) mínimo de ϕ ocurre en $\widehat{\boldsymbol{x}} = \boldsymbol{A}^{-1}\boldsymbol{b}$ y es dado por $-\frac{1}{2}\boldsymbol{b}^\top \boldsymbol{A}^{-1}\boldsymbol{b}$.

El mínimo puede ser encontrado de forma iterativa, usando el siguiente esquema:

$$\boldsymbol{x}^{(k+1)} = \boldsymbol{x}^{(k)} + \lambda^{(k)}\boldsymbol{p}^{(k)}, \qquad r = 0, 1, \dots, \tag{6}$$

donde $\lambda^{(k)}$ es un escalar (que representa un largo de paso), mientras que $\boldsymbol{p}^{(k)}$ es un vector que indica la dirección de búsqueda.

Método de gradientes conjugados

Resolver el sistema lineal $\boldsymbol{Ax} = \boldsymbol{b}$ es equivalente a determinar el mínimo de la función:

$$\phi(\boldsymbol{x}) = \tfrac{1}{2}\boldsymbol{x}^\top \boldsymbol{Ax} - \boldsymbol{x}^\top \boldsymbol{b}.$$

En efecto, por hacer la derivada de ϕ igual a $\mathbf{0}$, vemos que el punto estacionario de ϕ ocurre en el punto $\boldsymbol{x}$ donde $\boldsymbol{Ax} = \boldsymbol{b}$.

- Este procedimiento asume que la matriz $\boldsymbol{A}$ es simétrica y definida positiva.
- En este caso, el (único) mínimo de ϕ ocurre en $\widehat{\boldsymbol{x}} = \boldsymbol{A}^{-1}\boldsymbol{b}$ y es dado por $-\tfrac{1}{2}\boldsymbol{b}^\top \boldsymbol{A}^{-1}\boldsymbol{b}$.

El mínimo puede ser encontrado de forma iterativa, usando el siguiente esquema:

$$\boldsymbol{x}^{(k+1)} = \boldsymbol{x}^{(k)} + \lambda^{(k)}\boldsymbol{p}^{(k)}, \qquad r = 0, 1, \dots, \tag{6}$$

donde $\lambda^{(k)}$ es un escalar (que representa un largo de paso), mientras que $\boldsymbol{p}^{(k)}$ es un vector que indica la dirección de búsqueda.

Note que

$$\mathsf{d}_x\,\phi(\boldsymbol{x}) = \tfrac{1}{2}(\mathsf{d}\boldsymbol{x})^\top \boldsymbol{A}\boldsymbol{x} - (\mathsf{d}\boldsymbol{x})^\top \boldsymbol{b} = (\mathsf{d}\boldsymbol{x})^\top (\boldsymbol{A}\boldsymbol{x} - \boldsymbol{b}),$$

es decir,

$$\frac{\partial \phi(\boldsymbol{x})}{\partial \boldsymbol{x}} = \boldsymbol{A}\boldsymbol{x} - \boldsymbol{b}, \qquad -\frac{\partial \phi(\boldsymbol{x})}{\partial \boldsymbol{x}} := \boldsymbol{r}$$

es un residuo.

Definición 1

Un conjunto de vectores $\{\boldsymbol{z}_0, \boldsymbol{z}_1, \dots, \boldsymbol{z}_k\}$ se dicen conjugados[5] con respecto a la matriz simétrica y definida positiva $\boldsymbol{A}$ si

$$\boldsymbol{z}_i^\top \boldsymbol{A}\boldsymbol{z}_j = 0, \qquad \forall\, i \neq j.$$

Un conjunto de vectores satisfaciendo esta propiedad también son linealmente independientes.

[5] también son dichos $\boldsymbol{A}$-conjugados.

Ahora, note que

$$\boldsymbol{x}^{(k)} = \boldsymbol{x}^{(0)} + \lambda^{(1)}\boldsymbol{p}^{(1)} + \cdots + \lambda^{(k-1)}\boldsymbol{p}^{(k-1)}.$$

Escogiendo $\lambda^{(k)}$ como

$$\lambda^{(k)} = \frac{\boldsymbol{r}^{(k)\top}\boldsymbol{p}^{(k)}}{\boldsymbol{p}^{(k)\top}\boldsymbol{A}\boldsymbol{p}^{(k)}},$$

obtenemos que la dirección de búsqueda $\boldsymbol{p}^{(k)}$ es $\boldsymbol{A}$-conjugada con $\boldsymbol{p}^{(1)}, \dots, \boldsymbol{p}^{(k-1)}$.

- La secuencia $\{\boldsymbol{x}^{(k)}\}$ definida por por el algoritmo gradiente conjugado (6) converge a la solución $\widehat{\boldsymbol{x}}$ en a lo más n pasos.
- El algoritmo puede ser implementado muy eficientemente y por tanto es una opción atractiva para problemas de gran tamaño.

Algoritmo 1: Gradientes conjugados.

Entrada : Estimación inicial $\boldsymbol{x}^{(0)}$

Parámetros: Tolerancia τ.

begin

$\quad k \leftarrow 0$

$\quad$ Calcular $\boldsymbol{r}^{(k)} = \boldsymbol{b} - \boldsymbol{A}\boldsymbol{x}^{(k)}$, $\boldsymbol{p}^{(k)} = \boldsymbol{r}^{(k)}$ y $\gamma^{(k)} = \|\boldsymbol{r}^{(k)}\|^2$

$\quad$ **while** $\gamma^{(k)} > \tau$ **do**

$$\lambda^{(k)} = \frac{\boldsymbol{r}^{(k)\top}\boldsymbol{p}^{(k)}}{\boldsymbol{p}^{(k)\top}\boldsymbol{A}\boldsymbol{p}^{(k)}}$$

$$\boldsymbol{x}^{(k+1)} = \boldsymbol{x}^{(k)} + \lambda^{(k)}\boldsymbol{p}^{(k)}$$

$$\boldsymbol{r}^{(k)} = \boldsymbol{b} - \boldsymbol{A}\boldsymbol{x}^{(k+1)}$$

$$\beta^{(k)} = \frac{\boldsymbol{r}^{(k+1)\top}\boldsymbol{p}^{(k)}}{\boldsymbol{p}^{(k)\top}\boldsymbol{A}\boldsymbol{p}^{(k)}}$$

$\quad\quad \boldsymbol{p}^{(k+1)} = \boldsymbol{r}^{(k+1)} - \beta^{(k)}\boldsymbol{p}^{(k)}$ y $\gamma^{(k)} = \|\boldsymbol{r}^{(k)}\|^2$

$\quad\quad$ Hacer $k \leftarrow k + 1$ y $\boldsymbol{r}^{(k)} \leftarrow \boldsymbol{r}^{(k+1)}$, $\boldsymbol{p}^{(k)} \leftarrow \boldsymbol{p}^{(k+1)}$

$\quad$ **end**

$\quad$ **return** $\widehat{\boldsymbol{x}} = \boldsymbol{x}^{(k)}$

end