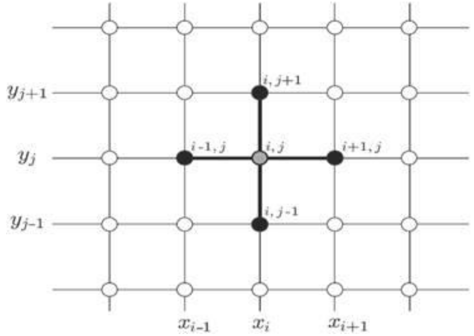
Métodos numéricos

Pablo Dmitruk

2do cuatrimestre 2021 - DF/FCEN/UBA

Método de Jacobi: $\phi_{i,j}^{n+1} = \frac{1}{4}(\phi_{i+1,j}^{n} + \phi_{i-1,j}^{n} + \phi_{i,j+1}^{n} + \phi_{i,j-1}^{n}) + \frac{\Delta^2}{4}\rho_{i,j}^{n}$



En general, $\mathbf{A}\,\bar{\mathbf{u}} = \bar{\mathbf{d}}$

$$\mathbf{A} = \begin{pmatrix} a_{11} & 0 & 0 & 0 & 0 \\ 0 & a_{22} & 0 & 0 & 0 \\ 0 & 0 & a_{33} & 0 & 0 \\ 0 & 0 & 0 & \dots & 0 \\ 0 & 0 & 0 & 0 & a_{mm} \end{pmatrix} + \begin{pmatrix} 0 & 0 & 0 & 0 & 0 \\ a_{21} & 0 & 0 & 0 & 0 \\ a_{31} & a_{32} & 0 & 0 & 0 \\ \dots & \dots & \dots & 0 & 0 \\ a_{m1} & a_{m2} & \dots & a_{m,m-1} & 0 \end{pmatrix} + \begin{pmatrix} 0 & a_{12} & a_{13} & \dots & a_{1m} \\ 0 & 0 & a_{23} & \dots & a_{2m} \\ 0 & 0 & 0 & \dots & \dots \\ 0 & 0 & 0 & 0 & a_{m-1,m} \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix}$$

D + **L** + **U**

$$\sum_{j=1}^{m} a_{ij} u_j = d_i \quad ; \quad i = 1, \ldots, m$$

$$u_i = \frac{1}{a_{ii}} \left(d_i - \sum_{j=1, j \neq i}^{m} a_{ij} u_j \right) \quad ; \quad i = 1, \ldots, m$$

$$u_i^{n+1} = \frac{1}{a_{ii}} \left(d_i^n - \sum_{j=1, j \neq i}^{m} a_{ij} u_j^n \right) \quad ; \quad i = 1, \ldots, m$$

iteración n → n+1

Forma matricial: $\mathbf{D}\bar{\mathbf{u}}^{n+1} = \bar{\mathbf{d}}^n - (\mathbf{L} + \mathbf{U})\bar{\mathbf{u}}^n$ → Jacobi

Se define el residuo $\bar{\mathbf{R}}^n = \mathbf{A}\bar{\mathbf{u}}^n - \bar{\mathbf{d}}^n = (\mathbf{D} + \mathbf{L} + \mathbf{U})\bar{\mathbf{u}}^n - \bar{\mathbf{d}}^n$

Si llamo $\Delta\bar{\mathbf{u}}^n = \bar{\mathbf{u}}^{n+1} - \bar{\mathbf{u}}^n \quad \rightarrow \quad \mathbf{D}\Delta\bar{\mathbf{u}}^n = \mathbf{D}\bar{\mathbf{u}}^{n+1} - \mathbf{D}\bar{\mathbf{u}}^n = \bar{\mathbf{d}}^n - (\mathbf{L} + \mathbf{U})\bar{\mathbf{u}}^n - \mathbf{D}\bar{\mathbf{u}}^n$

$\rightarrow \quad \mathbf{D}\Delta\bar{\mathbf{u}}^n = -\bar{\mathbf{R}}^n$ el método “conduce” $\Delta\bar{\mathbf{u}}^n$ a 0

Método de Gauss-Seidel (o de relajación)

→ uso valores ya iterados de $\phi^{n+1}_{i-1,j}$ y $\phi^{n+1}_{i,j-1}$ para obtener $\phi^{n+1}_{i,j}$

$$\phi^{n+1}_{i,j} = \frac{1}{4}(\phi^{n}_{i+1,j} + \phi^{n+1}_{i-1,j} + \phi^{n}_{i,j+1} + \phi^{n+1}_{i,j-1}) + \frac{\Delta^2}{4}\rho^{n}_{i,j}$$

→ acelero convergencia

→ se puede ahorrar memoria en un programa (porque no necesito guardar el paso anterior)

Forma general, $\mathbf{A}\,\bar{\mathbf{u}} = \bar{\mathbf{d}}$ $\qquad \sum_{j=1}^{m} a_{ij} u_j = d_i \quad ; \quad i = 1, \ldots, m$

$$u_i = \frac{1}{a_{ii}} \left(d_i - \sum_{j=1, j\neq i}^{m} a_{ij} u_j \right) \quad ; \quad i = 1, \ldots, m$$

$$u_i^{n+1} = \frac{1}{a_{ii}} \left(d_i^n - \sum_{j=1}^{i-1} a_{ij} u_j^{n+1} - \sum_{j=i+1}^{m} a_{ij} u_j^n \right) \quad ; \quad i = 1, \ldots, m$$

$$\mathbf{D}\bar{\mathbf{u}}^{n+1} + \mathbf{L}\bar{\mathbf{u}}^{n+1} = \bar{\mathbf{d}}^n - \mathbf{U}\bar{\mathbf{u}}^n$$

$$(\mathbf{D} + \mathbf{L})\bar{\mathbf{u}}^{n+1} = \bar{\mathbf{d}}^n - \mathbf{U}\bar{\mathbf{u}}^n$$

$$(\mathbf{D} + \mathbf{L})\Delta\bar{\mathbf{u}}^n = \bar{\mathbf{d}}^n - \mathbf{U}\bar{\mathbf{u}}^n - (\mathbf{D} + \mathbf{L})\bar{\mathbf{u}}^n = \bar{\mathbf{d}}^n - \mathbf{A}\bar{\mathbf{u}}^n = -\bar{\mathbf{R}}^n$$

→ se puede interpretar como otro método que “conduce” $\boldsymbol{\Delta}\bar{\mathbf{u}}^n$ a 0

Métodos sobre-relajación

Idea: incrementar la tasa de convergencia “propagando” las correcciones $\Delta\bar{\mathbf{u}}^n$ más rápido sobre la malla

Si $\tilde{\bar{\mathbf{u}}}^{n+1}$ es el valor obtenido en el esquema iterativo básico, el valor siguiente se obtiene cómo:

$$\bar{\mathbf{u}}^{n+1} = w\,\tilde{\bar{\mathbf{u}}}^{n+1} + (1-w)\,\bar{\mathbf{u}}^n = w\,(\tilde{\bar{\mathbf{u}}}^{n+1} - \bar{\mathbf{u}}^n) + \bar{\mathbf{u}}^n$$

$w =$ coeficiente de sobre-relajación

Jacobi + sobre-relajación

$$\phi_{i,j}^{n+1} = \frac{w}{4}(\phi_{i+1,j}^{n} + \phi_{i-1,j}^{n} + \phi_{i,j+1}^{n} + \phi_{i,j-1}^{n} + \Delta^2 \rho_{i,j}^{n}) + (1-w)\,\phi_{i,j}^{n}$$

Forma matricial: $\mathbf{A}\,\bar{\mathbf{u}} = \bar{\mathbf{d}}$

$$\mathbf{D}\,\bar{\mathbf{u}}^{n+1} = w\,\bar{\mathbf{d}}^{n} - w\,(\mathbf{L}+\mathbf{U})\,\bar{\mathbf{u}}^{n} + (1-w)\,\mathbf{D}\,\bar{\mathbf{u}}^{n}$$

$$\mathbf{D}\,\bar{\mathbf{u}}^{n+1} = w\,(\bar{\mathbf{d}}^{n} - \mathbf{A}\bar{\mathbf{u}}^{n}) + \mathbf{D}\,\bar{\mathbf{u}}^{n}$$

$-\bar{\mathbf{R}}^{n}$

$$\mathbf{D}\,\Delta\bar{\mathbf{u}}^{n} = -w\,\bar{\mathbf{R}}^{n}$$

Gauss-Seidel + sobre-relajación = SOR

(sobre-relajación sucesiva)

$$\tilde{\phi_{i,j}}^{n+1} = \frac{1}{4}(\phi^n_{i+1,j} + \phi^{n+1}_{i-1,j} + \phi^n_{i,j+1} + \phi^{n+1}_{i,j-1}) + \frac{\Delta^2}{4}\rho^n_{i,j}$$

$$\phi^{n+1}_{i,j} = w\,\tilde{\phi_{i,j}}^{n+1} + (1-w)\,\phi^n_{i,j} = w(\tilde{\phi_{i,j}}^{n+1} - \phi^n_{i,j}) + \phi^n_{i,j}$$

Forma matricial: $\mathbf{A}\,\bar{\mathbf{u}} = \bar{\mathbf{d}}$

$$\mathbf{D}\,\bar{\mathbf{u}}^{n+1} + w\,\mathbf{L}\bar{\mathbf{u}}^{n+1} = w\,\bar{\mathbf{d}}^n - w\,\mathbf{U}\,\bar{\mathbf{u}}^n + (1-w)\,\mathbf{D}\,\bar{\mathbf{u}}^n$$

$$(\mathbf{D} + w\,\mathbf{L})\,\Delta\bar{\mathbf{u}}^n = -w\,\bar{\mathbf{R}}^n$$

Cómo elegimos w ?

Convergencia de esquemas iterativos

Sup. $\bar{\mathbf{u_e}}$ sol. exacta de $\mathbf{A}\,\bar{\mathbf{u}} = \bar{\mathbf{d}} \iff \mathbf{A}\,\bar{\mathbf{u_e}} = \bar{\mathbf{d}}$

→ definimos el error en la iteración n como $\bar{\epsilon}^n = \bar{\mathbf{u}}^n - \bar{\mathbf{u_e}}$

sup. una subdivisión (splitting) arbitraria de $\mathbf{A} = \mathbf{P} + \mathbf{S}$

y un esquema iterativo $\mathbf{P}\,\bar{\mathbf{u}}^{n+1} = \mathbf{d} - \mathbf{S}\,\bar{\mathbf{u}}^n \iff \mathbf{P}\,\Delta\bar{\mathbf{u}}^n = -\bar{\mathbf{R}}^n = \bar{\mathbf{d}}^n - \mathbf{A}\,\bar{\mathbf{u}}^n$

$\mathbf{P} =$ matriz de convergencia o pre-condicionador

→ la elegimos de manera que sea fácil de invertir y “parecida” a $\mathbf{A}$

Notemos que $\mathbf{P}\,\bar{\epsilon}^{n+1} = \mathbf{P}\,\bar{\mathbf{u}}^{n+1} - \mathbf{P}\,\bar{\mathbf{u_e}} = \bar{\mathbf{d}}^n - \mathbf{S}\,\bar{\mathbf{u}}^n - \mathbf{P}\,\bar{\mathbf{u_e}} = -\mathbf{S}\,\bar{\epsilon}^n$

ya que $\mathbf{A}\,\bar{\mathbf{u_e}} = (\mathbf{P} + \mathbf{S})\,\bar{\mathbf{u_e}} = \bar{\mathbf{d}} \Rightarrow \mathbf{P}\,\bar{\mathbf{u_e}} = \bar{\mathbf{d}} - \mathbf{S}\,\bar{\mathbf{u_e}}$

$$\Rightarrow \bar{\epsilon}^{n+1} = -\mathbf{P}^{-1}\mathbf{S}\,\bar{\epsilon}^n = (\mathbf{I} - \mathbf{P}^{-1}\mathbf{A})\,\bar{\epsilon}^n = (\mathbf{I} - \mathbf{P}^{-1}\mathbf{A})^n\,\bar{\epsilon}^1$$

$$\mathbf{P}^{-1}\mathbf{A} = \mathbf{I} + \mathbf{P}^{-1}\mathbf{S}$$

$$\bar{\epsilon}^{n+1} = \mathbf{G}^n \; \bar{\epsilon}^1$$

$\mathbf{G} = \mathbf{I} - \mathbf{P}^{-1}\mathbf{A} =$ matriz de amplificación

Para que $|\bar{\epsilon}^{n+1}| \to 0$ entonces $\sigma(\mathbf{G}) \leq 1$

con $\sigma(\mathbf{G}) = \max(|\lambda_j(\mathbf{G})|)$ = radio espectral

autovalores

Notemos que $\bar{\mathbf{R}}^n = \mathbf{A}\,\bar{\mathbf{u}}^n - \bar{\mathbf{d}} = (\mathbf{A}\bar{\mathbf{u}}^n - \bar{\mathbf{d}}) - (\mathbf{A}\bar{\mathbf{u}}_{\mathbf{e}} - \bar{\mathbf{d}}) = \mathbf{A}\,(\bar{\mathbf{u}}^n - \bar{\mathbf{u}}_{\mathbf{e}}) = \mathbf{A}\,\bar{\epsilon}^n$

Si $\bar{\mathbf{R}}^n \to 0 \quad \Rightarrow \quad \bar{\epsilon}^n \to 0$

$\mathbf{G} = \mathbf{I} - \mathbf{P}^{-1}\mathbf{A}$ → depende de **P**

Jacobi: $\mathbf{D}\,\bar{\mathbf{u}}^{n+1} = \mathbf{d} - (\mathbf{L}+\mathbf{U})\,\bar{\mathbf{u}}^{n} \quad \rightarrow \mathbf{P} = \mathbf{D}\ ,\ \mathbf{S} = \mathbf{L}+\mathbf{U}$ $\qquad \mathbf{G} = \mathbf{I} - \mathbf{D}^{-1}(\mathbf{D}+\mathbf{L}+\mathbf{U}) = -\mathbf{D}^{-1}(\mathbf{L}+\mathbf{U})$

Gauss-Seidel: $(\mathbf{D}+\mathbf{L})\,\bar{\mathbf{u}}^{n+1} = \mathbf{d} - \mathbf{U}\,\bar{\mathbf{u}}^{n} \quad \rightarrow \mathbf{P} = \mathbf{D}+\mathbf{L}\ ,\ \mathbf{S} = \mathbf{U}$ $\qquad \mathbf{G} = \mathbf{I} - (\mathbf{D}+\mathbf{L})^{-1}(\mathbf{D}+\mathbf{L}+\mathbf{U}) = -(\mathbf{D}+\mathbf{L})^{-1}\mathbf{U}$

→ se puede ver que convergen si **A** es "*dominante diagonal*"

tasa de reducción de error $\qquad \Delta\epsilon^{n} = \left[\frac{|\epsilon^{n}|}{|\epsilon^{1}|}\right]^{1/n}$

Cómo $\quad \epsilon^{n} = (\mathbf{G})^{n}\epsilon^{1} \quad \Rightarrow \quad \Delta\epsilon^{n} \leq \sigma(\mathbf{G})$

S = tasa de convergencia = $|\log\sigma(\mathbf{G})|$

El error se reduce un factor 10 en un número de iteraciones n tal que

$$\left(\frac{1}{10}\right)^{1/n} \geq \sigma(G) \rightarrow 10^{-1/n} \geq \sigma(G) \rightarrow -\frac{1}{n} \geq \log\sigma \rightarrow -\log\sigma \geq \frac{1}{n} \rightarrow n \geq -\frac{1}{\log\sigma}$$

Ejemplo: ec. Poisson + met. Jacobi, c.c. periódicas

$$\mathbf{A} = \begin{pmatrix} -4 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & -4 & 1 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & -4 & 1 & 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 1 & -4 & 1 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 1 & -4 & 1 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 1 & -4 & 1 & 0 & 1 \\ 0 & 0 & 0 & 1 & 0 & 1 & -4 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 1 & -4 & 1 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 & -4 \end{pmatrix}$$

Se puede ver que $\sin\left(\frac{\pi\, l\, i}{M}\right)\, \sin\left(\frac{\pi\, m\, j}{M}\right)$ son autofunciones de **A**, con autovalor

$$\Omega = -4\left[\sin^2\left(\frac{\pi\, l\, i}{2M}\right) + \sin^2\left(\frac{\pi\, m\, j}{2M}\right)\right] \quad l, m = 1, \ldots, M$$

Como $\mathbf{G} = \mathbf{I} - \mathbf{P}^{-1}\mathbf{A} \quad \rightarrow \lambda = 1 - \frac{\Omega}{d}$ son autovalores de **G**

$$\lambda = 1 - \left[\sin^2\left(\frac{\pi\, l\, i}{2M}\right) + \sin^2\left(\frac{\pi\, m\, j}{2M}\right)\right] = \frac{1}{2}\left[\cos\left(\frac{\pi\, l}{M}\right) + \cos\left(\frac{\pi\, m}{M}\right)\right]$$

el autovalor más grande es $\sigma = \cos\frac{\pi}{M} < 1$

Si $M \gg 1 \;\rightarrow\; \sigma \simeq 1 - \frac{\pi^2}{2M^2}$

Para $M = 100 \;\rightarrow \sigma = 0.9995$, $\log\sigma = -0.0002 \;\; -1/\log\sigma \simeq 4600$

Reduzco el error un factor 10 en 4600 iteraciones

para Gauss-Seidel, $\mathbf{G} = \mathbf{I} - (\mathbf{D} + \mathbf{L})^{-1}\,\mathbf{A}$

y se puede ver que $\lambda = \frac{1}{4}\left[\cos\left(\frac{\pi\, l}{M}\right) + \cos\left(\frac{\pi\, m}{M}\right)\right]^2$

y el valor máximo es $\sigma = \cos^2\frac{\pi}{M} \simeq 1 - \frac{\pi^2}{M^2}$, $M \gg 1$

→ converge más rápido que Jacobi

Para $M = 100 \rightarrow -1/\log\sigma \simeq 2331$

Para el método SOR vimos que $(\mathbf{D} + w\,\mathbf{L})\,\Delta\bar{\mathbf{u}}^n = -w\,\bar{\mathbf{R}}^n$

y de la definición de $\bar{\epsilon}^n = \bar{\mathbf{u}}^n - \bar{\mathbf{u}_e}$ $\quad \mathbf{A}\,\bar{\mathbf{u}_e} = \bar{\mathbf{d}}$

se obtiene, $\bar{\epsilon}^{n+1} = \mathbf{G}\,\bar{\epsilon}^n$ $\quad \mathbf{G} = (\mathbf{D} + w\,\mathbf{L})^{-1}[(1-w)\mathbf{D} - w\mathbf{U}]$

es difícil sacar los autovalores de G pero se puede usar que $\det(\mathbf{G}) = \prod_{j=1}^{M} \lambda_j$

$$\text{Si } \det(\mathbf{G})^M \leq 1 \;\Rightarrow\; \sigma(\mathbf{G}) \leq 1$$

usando que son matrices triangulares se puede ver que $\det(\mathbf{G}) = (1-w)^M$

$$\rightarrow \sigma^M \geq (1-w)^M \;\Rightarrow\; |1-w| \leq \sigma < 1 \;\Rightarrow\; 0 < w < 2$$

Existe un valor w_{opt} que minimiza $\sigma(G)$

$$\rightarrow w_{opt} = \frac{2}{1 + \sqrt{1 - \sigma_J^2}}$$

y en ese caso $\sigma(w_{opt}) = w_{opt} - 1$

σ_J = máximo autovalor c/ met. Jacobi

Para el problema de Poisson sabemos que $\sigma_J = \cos\frac{\pi}{M}$

$$\rightarrow w_{opt} = \frac{2}{1+\sin\frac{\pi}{M}} \simeq 2\left(1 - \frac{\pi}{M} + \frac{\pi^2}{M^2}\right) \quad , \quad M \gg 1$$

$$\rightarrow \sigma_{SOR} \simeq 1 - \frac{2\pi}{M} + O\left(\frac{1}{M^2}\right)$$

que es mucho más chico que el sigma de Gauss-Seidel o Jacobi

$$\text{Para } M = 100 \rightarrow -1/\log\sigma \simeq 35 \qquad !!$$

Otros métodos iterativos → Gradiente Conjugado y GMRES

CGM

Generalized Minimal Residual Method

Para matrices simétricas y def. positiva.
Des. en 1952, Hestenes & Stiefel

Matrices no simétricas.
Des. en 1986, Saad & Schultz