Mètodes Numèrics II

Joan Carles Tatjer
Departament de Matemàtiques i Informàtica
Universitat de Barcelona
Gran Via 585, 08007 Barcelona, Spain
E-mail: jcarles@maia.ub.es

2 de novembre de 2020

Capítol 1

Mètodes iteratius de resolució de sistemes lineals

1.1 Introducció

Recordem que un mètode directe és un mètode que, suposant que totes les operacions aritmètiques són exactes, troba la solució exacta d'un sistema lineal compatible i determinat $Ax = b$ amb un nombre finit de passos (operacions). L'exemple més simple és el mètode de Gauss, en el que el sistema inicial es transforma en un sistema triangular. Els mètodes directes de resolució de sistemes lineals son costosos si la dimensió de la matriu corresponent és gran (recordem que cal fer de l'ordre de n^3 operacions, on n és la dimensió del sistema). Finalment, podem tenir també problemes de memòria, si la dimensió n és molt gran, tot i que no tingui gaires elements diferents de zero.

Nosaltres considerarem mètodes iteratius en els que busquem la solució intentant usar poques operacions per iteració. Si la convergència és suficientment ràpida, el procediment acabarà amb una bona aproximació i sense masses operacions. Aquest pot ser el cas, si la matriu del sistema té pocs elements diferents de zero combinat amb una dimensió elevada de la matriu. Abans d'entrar en matèria, recordarem algunes propietats de les normes en espais vectorials.

1.1.1 Normes vectorials i normes matricials

Sigui E un espai vectorial sobre $\mathbb{R}$ o $\mathbb{C}$.

Definició 1.1.1 *Una **norma** a E és una aplicació*

$$\begin{array}{rccc} \|\cdot\| & : & E & \to & \mathbb{R}^+ \\ & & x & \mapsto & \|x\| \end{array}$$

complint:

a) $\|x\| = 0$ *sii* $x = 0$.

b) $\|cx\| = |c|\|x\|$, *per a tot* $x \in E$ *i tot escalar (real o complex).*

c) $\|x + y\| \leq \|x\| + \|y\|$, *per a tot* $x, y \in E$ *(desigualtat triangular),*

A la parella $(E, \|\cdot\|)$ *l'anomenem* ***espai normat****.*

Els espais normats més comuns són $E = \mathbb{R}^n$ i $E = \mathbb{C}^n$. Les normes sobre aquests espais les anomenarem normes vectorials. Si ens volem referir indistintament a $\mathbb{R}^n$ o $\mathbb{C}^n$, parlarem de l'espai $\mathbb{K}^n$, on $\mathbb{K} = \mathbb{R}$ o $\mathbb{K} = \mathbb{C}$. Les més emprades són les normes L_p o normes de Hölder:

Si $x \in \mathbb{K}^n$ $x = (x_1, \ldots, x_n)$, definim

$$\|x\|_p = [|x_1|^p + |x_2|^p + \cdots + |x_n|^p]^{1/p}, \qquad 1 \leq p < \infty.$$

Els casos particulars més importants són la **norma euclidiana** ($p = 2$), la **norma sub-1** ($p = 1$). Un altre cas important s'obtè fent $p \to \infty$, amb el que obtenim la **norma del màxim**, també anomenada **norma sub-infinit**:

$$\|x\|_\infty = \max_{1 \leq i \leq n} |x_i|.$$

Considerem ara l'espai vectorial $\mathbb{K}^{n \times n}$ format per les matrius $n \times n$ sobre $\mathbb{K}$.

Definició 1.1.2 *Una* ***norma matricial*** *(o sub-multiplicativa) és una norma en l'espai vectorial* $\mathbb{K}^{n \times n}$ *que, a més, és sub-multiplicativa:*

$$\|AB\| \leq \|A\| \, \|B\|, \qquad \forall\, A, B \in \mathbb{K}^{n \times n}.$$

És fàcil demostrar que per a tota norma matricial $\|I\| \geq 1$.

Nota 1.1.1 *No totes les normes són matricials Per exemple, si definim, per a*

$$A = \begin{pmatrix} a & b \\ c & d \end{pmatrix},$$

$\|A\| = \max\{|a|, |b|, |c|, |d|\}$*, tenim que si*

$$A = \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix},$$

llavors $\|A\| = 1$*, i* $\|AA\| = 2$*.*

Definició 1.1.3 *Una norma matricial* $\|\cdot\|$ *és* ***consistent*** *(o compatible) amb una norma vectorial (que notarem igual) sii*

$$\|Ax\| \leq \|A\| \, \|x\|, \qquad \forall\, A \in \mathbb{K}^{n \times n}, \quad \forall\, x \in \mathbb{K}^n.$$

Més en general, donades tres normes, $\|\cdot\|_1$ *de* $\mathbb{K}^n$*,* $\|\cdot\|_2$ *de* $\mathbb{K}^m$ *i* $\|\cdot\|_3$ *de* $\mathbb{K}^{m \times n}$*, diem que són consistents si* $\|Ax\|_2 \leq \|A\|_3 \, \|x\|_1$*.*

Nota 1.1.2 *La noció de consistència es pot generalitzar de la següent manera: Diem que les normes* f_1*,* f_2 *i* f_3 *en* $\mathbb{K}^{m \times q}$*,* $\mathbb{K}^{m \times n}$ *i* $\mathbb{K}^{n \times q}$ *són* ***mutuament consistents*** *si per totes les matrius* $A \in \mathbb{K}^{m \times n}$ *i* $B \in \mathbb{K}^{n \times q}$ *tenim* $f_1(AB) \leq f_2(A) f_3(B)$*.*

Donada una norma vectorial es pot construir una norma matricial consistent amb ella:

Proposició 1.1.1 *Sigui $\|\cdot\|_1$ una norma vectorial en $\mathbb{K}^n$, $\|\cdot\|_2$ en $\mathbb{K}^m$. Per a tota matriu $A \in \mathbb{K}^{m\times n}$ definim*

$$\|A\| = \sup_{x\neq 0} \frac{\|Ax\|_2}{\|x\|_1}.$$

*Llavors aquesta norma és una norma matricial consistent amb la norma vectorial, que anomenen norma matricial **induïda** o natural.*

Demostració:
Per per simplificar notació usarem $\|\cdot\|$ per a les tres normes. En primer lloc, notem que

$$\frac{\|Ax\|}{\|x\|} = \left\| \frac{1}{\|x\|} Ax \right\| = \left\| A\left(\frac{x}{\|x\|}\right) \right\|.$$

Per tant,

$$\sup_{x\neq 0} \frac{\|Ax\|}{\|x\|} = \sup_{\|y\|=1} \|Ay\| = \max_{\|y\|=1} \|Ay\| < \infty,$$

és a dir, la norma està ben definida. Veiem que és norma:

- $\|A\| = 0$ sii $A = 0$.

  És obvi que si $A = 0$ llavors $\|A\| = 0$. Suposem que $\|A\| = 0$. Llavors $\max_{\|y\|=1} \|Ay\| = 0$. Per tant, $\|Ay\| = 0$, per a tot y tal que $\|y\| = 1$, i per tant per a tot $x \neq 0$

  $$\|Ax\| = \|x\| \left\| A\left(\frac{x}{\|x\|}\right) \right\| = 0.$$

  Això implica que $Ax = 0$, per a tot x, i per tant, $A = 0$.

- $\|cA\| = |c|\|A\|$.

  $$\|cA\| = \max_{\|y\|=1} \|(cA)y\| = |c| \max_{\|y\|=1} \|Ay\| = |c|\|A\|.$$

- $\|A + B\| \leq \|A\| + \|B\|$.

  $$\|A + B\| = \max_{\|y\|=1} \|(A + B)y\| = \max_{\|y\|=1} \|Ay + By\| \leq$$

  $$\max_{\|y\|=1}(\|Ay\| + \|By\|) \leq \max_{\|y\|=1} \|Ay\| + \max_{\|y\|=1} \|By\| = \|A\| + \|B\|.$$

És fàcil veure que la norma és subordinada. Si $x \neq 0$:

$$\|Ax\| = \left\| A\frac{x}{\|x\|} \right\| \|x\| \leq \max_{\|y\|=1} \|Ay\| \, \|x\| = \|A\| \, \|x\|.$$

i que és multiplicativa:

$$\|AB\| = \max_{\|y\|=1} \|(AB)y\| = \max_{y\|=1} \|A(By)\| \leq \max_{\|y\|=1} \|A\| \, \|By\| = \|A\| \, \|B\|.$$

□

Nota 1.1.3 *Donada una norma $\|\cdot\|$ de $\mathbb{K}^n$, i un vector $x \in \mathbb{K}^n$, podem identificar x amb una matriu de $\mathbb{K}^{n\times 1}$. Llavors, la norma matricial induïda de x com a matriu, coincideix amb la norma de x com a vector, si agafem com a norma en $\mathbb{K}$ el valor absolut si $\mathbb{K}=\mathbb{R}$ o el mòdul si $\mathbb{K}=\mathbb{C}$.*

Proposició 1.1.2 *Sigui $A=(a_{ij})_{1\le i\le n,1\le j\le m}\in\mathbb{R}^{n\times m}$. Aleshores*

$$\|A\|_1 = \max_{1\le j\le m}\left(\sum_{i=1}^{n}|a_{ij}|\right),$$
$$\|A\|_2 = \left(\rho\left(A^\top A\right)\right)^{1/2} \qquad \textit{(norma euclidiana)},$$
$$\|A\|_\infty = \max_{1\le i\le n}\left(\sum_{j=1}^{m}|a_{ij}|\right) \qquad \textit{(norma del màxim)},$$

on $\rho(M)$ designa el radi espectral, màxim dels mòduls dels valors propis d'una matriu M.

Demostració:
Només demostrarem el cas de la norma euclidana. Per això necessitem els següents resultats:

Proposició 1.1.3 *Sigui $M\in\mathbb{R}^{m\times m}$ una matriu real i simètrica ($M^T=M$). Llavors* $\mathrm{Spec}(M)\subset\mathbb{R}$.

Demostració:
Si $\lambda\in\mathrm{Spec}(M)$ llavors existeix $v\in\mathbb{C}^n$, $v\neq 0$, tal que $Mv=\lambda v$. Com que $M^T=M$ tenim que $\overline{v}^TM=\overline{\lambda}\overline{v}^T$. Per tant,

$$\overline{v}^TMv=\lambda\overline{v}^Tv, \qquad \overline{v}^TMv=\overline{\lambda}\overline{v}^Tv.$$

Com que $v\neq 0$, això implica que $\overline{\lambda}=\lambda$, i per tant, $\lambda\in\mathbb{R}$. □.

Proposició 1.1.4 *Tota matriu real i simètrica diagonalitza en una base ortonormal.*

Aquest resultat el demostrarem més endavant.

Demostració de la Proposició 1.1.2:
La matriu $M=A^TA\in\mathbb{R}^{m\times m}$ és simètrica i semidefinida positiva ($x^TA^TAx=\|Ax\|_2\ge 0$). Per tant, diagonalitza en una base ortonormal, $v_1,\ldots,v_m$ i els seus valors propis són reals no negatius:

$$\lambda_1\ge\lambda_2\ge\cdots\ge\lambda_m\ge 0.$$

Per definició $\|A\|_2=\max_{\|x\|_2=1}\|Ax\|_2$. Per tant,

$$\|A\|_2^2=\max_{\|x\|_2=1}\|Ax\|^2=\max_{\|x\|_2=1}x^TA^TAx.$$

Donat un vector x, podem escriure

$$x=\alpha_1v_1+\ldots+\alpha_mv_m.$$

Llavors:

$$\|A\|_2^2=\max_{\|x\|_2=1}(\alpha_1v_1+\ldots+\alpha_mv_m)^T(\alpha_1\lambda_1v_1+\ldots+\alpha_m\lambda_nv_m)=\max_{\|x\|_2=1}\alpha_1^2\lambda_1+\cdots+\alpha_m^2\lambda_m\le\lambda_1.$$

Aquesta darrera desigualtat és deguda a que $\alpha_1^2 + \cdots + \alpha_m^2 = 1$ i

$$\alpha_1^2\lambda_1 + \cdots + \alpha_m^2\lambda_m \leq (\alpha_1^2 + \cdots + \alpha_m^2)\lambda_1 = \lambda_1.$$

Com que si agafem $x = v_1$ tenim que $v_1^T A^T A v = \lambda_1$, deduïm que $\|A\|_2^2 = \lambda_1 = \rho(A^T A)$. □

Podem donar una demostració alternativa que no usa la proposició 1.1.4.
En primer lloc tenim el següent resultat:

Lema 1.1.1 *Sigui A una matriu $n \times n$ i $f(x) = x^T Ax$, on $x \in \mathbb{R}^n$. Aleshores $Df(x) = x^T(A + A^T)$.*

Demostració:
Sigui $h \in \mathbb{R}^n$. En general, sabem que $Df(x)h = g'(0)$, on $g(t) = f(x + th)$. En el nostre cas

$$g(t) = x^T Ax + th^T Ax + tx^T Ah + t^2 h^T Ah = x^T Ax + t(x^T A^T h + x^T Ah) + t^2 h^T Ah.$$

Per tant, $Df(x)h = g'(0) = x^T(A + A^T)h$, el que implica el que voliem. □

A continuació tenim:

Proposició 1.1.5 *Si A és una matriu $n \times m$ llavors existeix un vector $z \in \mathbb{R}^m$ unitari tal que $A^T Az = \mu^2 z$, on $\mu = \|A\|_2$.*

Demostració:
Sigui z un vector unitari tal que $\|Az\|_2 = \|A\|_2$. Aleshores usant el teorema dels multiplicadors de Lagrange, existeix $\lambda \in \mathbb{R}$ tal que $z^T A^T A - \lambda z^T = 0$, o $A^T Az = \lambda z$. Com que $z^T z = 1$,tenim que

$$\mu^2 = \|A\|_2^2 = \|Az\|^2 = z^T A^T Az = \lambda,$$

amb el que hem demostrat la proposició. □

Per acabar la demostració, sabem que els valors propis de $A^T A$ són reals no negatius i si λ és un valor propi de $A^T A$ llavors existeix $x \in \mathbb{R}^m$ tal que $x^T x = 1$ i $A^T Ax = \lambda x$. Per tant, $\|Ax\|_2^2 = \lambda \leq \|A\|^2 = \mu^2$. Això implica que $\|A\|_2 = \sqrt{\rho(A^T A)}$.

Nota 1.1.4 *La proposició 1.1.2 és certa també per $\mathbb{C}^{n\times n}$ en els casos de les normes $\|\cdot\|_1$ i $\|\cdot\|_\infty$, tenint en compte que cal canviar el valor absolut pel mòdul. En el cas de la norma euclidiana, el resultat és lleugerament diferent: Donada una matriu $A \in \mathbb{C}^{n\times m}$ definim $A^H = \overline{A}^T$, on els elements de $\overline{A}$ són els conjugats dels elements de A, és a dir $\overline{A} = (\overline{a_{ij}})_{1\leq i\leq n, 1\leq j\leq m}$. Llavors, $\|A\|_2 = \sqrt{\rho(A^H A)}$.*

Acabarem amb tres resultats que relacionen el radi espectral d'una matriu amb una norma de la matriu induïda per una norma vectorial. Primer, però, necessitem enunciar el següent teorema:

Teorema 1.1.1 *Totes les normes a $\mathbb{K}^n$ són equivalents, és a dir, donades dues normes $\|\cdot\|$ i $\|\cdot\|'$ existeixen constants $k_1, k_2 > 0$ tals que*

$$k_1\|x\| \leq \|x\|' \leq k_2\|x\|, \qquad \forall\, x \in \mathbb{K}.$$

Nota 1.1.5 *Aquest teorema implica que si una successió $(x^{(k)})_{k\geq 0}$, on $x^{(k)} \in \mathbb{K}^n$ per a tot $k \geq 0$, és convergent per una norma, ho és per totes.*

Proposició 1.1.6 *Denotem per $\|\cdot\|$ una norma vectorial i la seva norma matricial induïda. Llavors, per a tota matriu $A \in \mathbb{K}^{n\times n}$ es verifica $\rho(A) \leq \|A\|$.*

Demostració:

En primer lloc considerarem el cas $\mathbb{K} = \mathbb{C}$:

Sigui λ el valor propi de mòdul màxim, i $v \in \mathbb{C}^n$ un vector propi no nul de valor propi $\lambda \in \mathbb{C}$. Podem suposar que és unitari: $Av = \lambda v$ i $\|v\| = 1$. Llavors,

$$|\lambda| = \|Av\| \leq \max_{\|x\|=1} \|Ax\| = \|A\|.$$

Per tant, $\rho(A) \leq \|A\|$ en aquest cas.

Considerem ara el cas $\mathbb{K} = \mathbb{R}$: Sigui $\|\cdot\|'$ la restricció d'una norma per $\mathbb{C}^n$ a $\mathbb{R}^n$. Obviament la seva restricció també és una norma a $\mathbb{R}^n$. Com que totes les normes són equivalents, existeix una constant $c > 0$ tal que

$$\|B\|' \leq c\|B\|,$$

per a tota matriu real $B \in \mathbb{R}^{n\times n}$. Ara podem aplicar el resultat pel cas complex. Tenim que per a tot $k \geq 1$:

$$\rho(A)^k = \rho(A^k) \leq \|A^k\|' \leq c\|A^k\| \leq c\|A\|^k.$$

Per tant,

$$\rho(A) \leq c^{1/k}\|A\|, \qquad \forall\, k \geq 1,$$

i com que $\lim_{k\to\infty} c^{1/k} = 1$, tenim el resultat que voliem també en aquest cas. □

Teorema 1.1.2 *Siguin $A \in \mathbb{K}^{n\times n}$ i $\epsilon > 0$. Llavors existeix una norma matricial induïda $\|\cdot\|_{A,\epsilon}$ tal que*

$$\|A\|_{A,\epsilon} \leq \rho(A) + \epsilon.$$

Nota 1.1.6 *Del teorema es dedueix que $\rho(A) = \inf_{\|\cdot\|} \|A\|$, però ρ no és una norma, ja que pot ser que $\rho(A) = 0$ i $A \neq 0$. Per exemple podem agafar qualsevol matriu triangular no nul·la amb la diagonal zero.*

Proposició 1.1.7 *Sigui $A \in \mathbb{K}^{n\times n}$ una matriu quadrada i $\|\cdot\|$ una norma consistent. Aleshores*

$$\lim_{k\to\infty} \|A^k\|^{1/k} = \rho(A).$$

1.2 Mètodes iteratius

1.2.1 Introducció

Suposem que volem resoldre per x el sistema $Ax = b$, on $A \in \mathbb{K}^{n\times n}$, $x, b \in \mathbb{K}^n$. La idea bàsica dels mètodes iteratius és construir una successió de vectors $(x^{(k)})_k$ tals que convergeix cap a la solució x. A la pràctica, el procès iteratiu es para quan $\|x^{(k)} - x\| < \epsilon$, on ϵ és una tolerància fixada i $\|\cdot\|$ és una norma vectorial. Com que la solució x no és coneguda, caldrà trobar criteris de parada calculables.

Un problema model

Considerem l'equació de Laplace en dimensió 2 amb condicions de contorn:

$$u_{xx} + u_{yy} = 0, \qquad (x, y) \in \Omega = (0, 1) \times (0, 1),$$

$$u(x, 0) = f(x), \quad u(x, 1) = g(x), \quad u(0, y) = \phi(y), \quad u(1, y) = \psi(y),$$

on f, g, ϕ, ψ són funcions donades i $(x, y) \in \overline{\Omega}$. Volem trobar la solució u del problema de contorn corresponent.

Trobarem una solució aproximada de la següent manera: considerem una malla de punts $(x_i, y_j) = (ih, jh)$, on $h = 1/(n + 1)$, $0 \leq i, j \leq n + 1$, i aproximem les derivades segones per quocients incrementals. Si u_{ij} és el valor aproximat de $u(x_i, x_j)$ llavors

$$u_{xx}(x_i, y_j) \approx \frac{u_{i+1,j} - 2u_{i,j} + u_{i-1,j}}{h^2},$$

$$u_{yy}(x_i, y_j) \approx \frac{u_{i,j+1} - 2u_{i,j} + u_{i,j-1}}{h^2},$$

i substituint les derivades parcials pels quocients incrementals en l'equació de Laplace, obtenim:

$$\frac{1}{h^2}(u_{i+1,j} + u_{i-1,j} + u_{i,j+1} + u_{i,j-1} - 4u_{ij}) = 0, \qquad 0 < i, j < n + 1.$$

A més, les condicions de contorn impliquen que

$$u_{i,0} = f_i, \quad u_{i,n} = g_i, \quad 0 \leq i \leq n, \qquad u_{0,j} = \varphi_j, \quad u_{n,j} = \psi_j, \quad 0 \leq j \leq n,$$

on $f_i = f(x_i)$, $g_i = g(x_i)$, $\varphi_j = \varphi(y_j)$, $\psi_j = \psi(y_j)$.

Si ordenem els punts de la xarxa fila per fila, obtenim el sistema

$$Av = b, \quad A \in \mathbb{R}^{n^2 \times n^2}, \quad v = (u_1, \ldots, u_n),$$

on $u_i = (u_{i,1}, \ldots, u_{i,n})$ i b es pot formar a partir dels valors de u a la frontera. A més:

$$A = \begin{pmatrix} 2I + T_n & -I & & \\ -I & 2I + T_n & \ddots & \\ & \ddots & \ddots & -I \\ & & -I & 2I + T_n \end{pmatrix},$$

$$T_n = \begin{pmatrix} 2 & -1 & & \\ -1 & 2 & \ddots & \\ & \ddots & \ddots & -1 \\ & & -1 & 2 \end{pmatrix}.$$

La matriu A és simètrica i definida positiva, i té una estructura de banda, amb subdiagonals buides a l'interior. Si fem descomposició de Cholesky de A, la corresponent matriu triangular inferior L serà una matriu tringular plena. Per tant, l'esforç de calcular aquesta descomposició és molt més

gran que la d'avaluar el producte de A per un vector. Veurem que en els mètodes iteratius podem treure avantage d'aquesta propietat.

La manera més simple per a generar una successió $(x^{(k)})_{k\geq 0}$ que convergeixi a la solució de $Ax = b$, on $A \in \mathbb{R}^{n\times n}$ és una matriu no singular i $b \in \mathbb{R}^n$, és considerar mètodes iteratius de la forma

$$\text{Donat } x^{(0)}, \quad x^{(k+1)} = Bx^{(k)} + c, \quad k \geq 0. \tag{1.1}$$

De manera natural, volem que si la successió $\{x^{(k)}\}_{k\geq 0}$ convergeix, $\lim_{k\to\infty} x^{(k)} = x^{(\infty)}$, llavors $x^{(\infty)}$ satisfà $Ax^{(\infty)} = b$. Notem que per continuïtat, $x^{\infty} = Bx^{\infty} + c$, i per tant caldrà que $c = (I - B)A^{-1}b$. Això dona peu a la següent definició:

Definició 1.2.1 *Diem que un mètode iteratiu de la forma (1.1) és* ***consistent*** *amb $Ax = b$ si $c = (I - B)A^{-1}b$.*

Definició 1.2.2 *Anomenem al mètode iteratiu (1.1),* ***Mètode iteratiu (lineal) estacionari (de primer ordre)****, i a la matriu B,* ***matriu d'iteració*** *del mètode.*

Nota 1.2.1 *El mètode iteratiu (1.1) és un cas especial de mètodes iteratius de la forma*

$$x^{(n+1)} = f_{n+1}(x^{(n)}, x^{(n-1)}, \ldots, x^{(n-m)}, A, b), \quad n \geq m,$$

on f_i i $x^{(m)}, \ldots, x^{(0)}$, són funcions i vectors donats, respectivament. Aquí $m+1$ és l'ordre del mètode. Si els f_i no depenen de i anomenem al mètode estacionari (en cas contrari no estacionari) i si f_i depèn linealment de $x^{(0)}, \ldots, x^{(m)}$ l'anomenem lineal (en cas contrari no lineal).

Definició 1.2.3 *Diem que un mètode iteratiu és* ***convergent*** *si la successió $(x_k^{(k)})_k$ corresponent convergeix per a tot vector inicial $x^{(0)}$.*

Nota 1.2.2 *Obviament, un mètode consistent no té perquè ser convergent. Per exemple, si $A = 2I$ llavors el mètode iteratiu $x^{(k+1)} = -x^{(k)} + b$ és consistent, però si $x^{(0)} = 0$ no és convergent.*

El nostre objectiu serà obtenir mètodes d'aquest tipus i demostrar en quines condicions convergeixen.

1.2.2 Convergència de mètodes iteratius estacionaris

Suposem que tenim un mètode iteratiu estacionari consistent (1.1) amb matriu d'iteració B. Si definim $e^{(k)} = x^{(k)} - \bar{x}$, on $A\bar{x} = b$, llavors el mètode és convergent sii $\lim_{k\to\infty} e^{(k)} = 0$, per a tot $x^{(0)}$. En primer lloc, podem donar una condició suficient de convergència, que ens serà molt útil:

Teorema 1.2.1 *Una condició suficient per a que un mètode iteratiu estacionari consistent $x^{(k+1)} = Bx^{(k)} + c$ sigui convergent és que $\|B\| < 1$, per alguna norma matricial consistent $\|\cdot\|$.*

Demostració:
Notem que $e^{(k+1)} = Bx^{(k)} + c - \bar{x} = Bx^{(k)} - B\bar{x} = Be^{(k)}$. Agafant normes:

$$\|e^{(k+1)}\| \leq \|B\| \, \|e^{(k)}\|,$$

i per tant

$$\|e^{(k)}\| \leq \|B\|^k \|e^{(0)}\|,$$

el que prova que $\|e^{(k)}\| \to 0$ quan $k \to \infty$, i per tant el teorema. □

Definició 1.2.4 *Diem que una matriu $B \in \mathcal{M}_{n\times n}$ és convergent, si $\lim_{k\to\infty} B^k = 0$.*

Teorema 1.2.2 *Sigui $B \in \mathbb{K}^{n\times n}$, i considerem un mètode estacionari consistent amb matriu d'iteració B. Les següents condicions són equivalents:*

a) B és convergent.

b) $\lim_{k\to\infty} B^k x = 0$, per a tot $x \in \mathbb{K}^n$.

c) $\rho(B) < 1$.

d) $\|B\| < 1$, per alguna norma matricial consistent.

e) El mètode iteratiu és convergent.

Demostració:
Com que b) i e) són trivialment equivalents, només caldrà demostrar a) $\Rightarrow$ b) $\Rightarrow$ c) $\Rightarrow$ d) $\Rightarrow$ a).

a) $\Rightarrow$ b) De la desigualtat $\|B^k x\| \leq \|B\|^k \|x\|$, que és certa per a qualsevol vector x, tenim que a) implica b).

b) $\Rightarrow$ c) Primer suposem que $\mathbb{K} = \mathbb{C}$. Si $\rho(B) \geq 1$ llavors existeix un valor propi $\lambda \in \mathbb{C}$ tal que $|\lambda| \geq 1$. Sigui $x \in \mathbb{C}^n$ un vector propi de valor propi λ. Aleshores la successió $B^k x = \lambda^k x$, $k = 1, 2, \ldots$ no convergeix a zero quan $k \to \infty$ i així b) implica c). En el cas que $\mathbb{K} = \mathbb{R}$, si existeix un valor propi real λ tal que $|\lambda| \geq 1$, podem procedir com abans. Si $\lambda \in \mathbb{C} \setminus \mathbb{R}$, sigui $x + iy$ un vector propi de valor propi λ. Com que la convergència no depèn de la norma escollida, treballarem amb la norma del suprem. Com abans, tenim que $B^k(x+iy)$ no convergeix a zero quan $k \to \infty$. A més,

$$\|B^k(x+iy)\|_\infty \leq \|B^k x\|_\infty + \|B^k y\|_\infty,$$

el que implica que $B^k x$ o $B^k y$ no convergeixen a zero quan $k \to \infty$. Per tant, també és certa la implicació en aquest cas.

c) $\Rightarrow$ d) Pel teorema 1.1.2, si $\rho(B) < 1$ existeix una norma matricial consistent tal que $\|B\| < 1$.

d) $\Rightarrow$ a) És conseqüència de que $\|B^k\| \leq \|B\|^k$.

□

Suposem que tenim un mètode iteratiu convergent amb matriu d'iteració B. Sabem que existeix una norma matricial consistent $\|\cdot\|$ tal que $\|B\| = \beta < 1$. En aquest cas, podem donar una estimació de l'error:

Proposició 1.2.1 *Sigui $\beta = \|B\| < 1$, i sigui $\bar{x}$ la solució del sistema $Ax = b$. Llavors*

$$\|x^{(k)} - \bar{x}\| \leq \frac{\beta}{1-\beta} \|x^{(k)} - x^{(k-1)}\|. \tag{1.2}$$

A més:

$$\|x^{(k)} - \bar{x}\| \leq \frac{\beta^k}{1-\beta} \|x_1 - x_0\|. \tag{1.3}$$

Demostració:

$$x^{(k)} - \bar{x} = B(x^{(k-1)} - \bar{x}) = B(x^{(k-1)} - x^{(k)}) + B(x^{(k)} - \bar{x}).$$

Agafant normes:

$$\|x^{(k)} - \bar{x}\| \leq \|B\|\|x^{(k)} - x^{(k-1)}\| + \|B\| \, \|x^{(k)} - \bar{x}\|,$$

i per tant,

$$(1 - \|B\|)\|x^{(k)} - \bar{x}\| \leq \|B\| \, \|x^{(k)} - x^{(k-1)}\|.$$

Finalment, com que $1 - \|B\| > 0$:

$$\|x^{(k)} - \bar{x}\| \leq \frac{\|B\|}{1 - \|B\|}\|x^{(k)} - x^{(k-1)}\|.$$

La segona part es dedueix immediatament de la primera. □

Nota 1.2.3 *Obviament, si $\beta \leq 1/2$ tenim que $\|x^{(k)} - \bar{x}\| \leq \|x^{(k)} - x^{(k-1)}\|$.*

Criteris de parada per a mètodes iteratius lineals estacionaris

En aquesta secció, volem indicar com estimar l'error en un mètode iteratiu convergent, amb el que tindrem un criteri de parada, és a dir, que cal imposar per a que l'error sigui menor que una certa tolerància, i com obtenir el nombre k_{min} d'iteracions necessàries per a reduïr l'error inicial per un factor ϵ.

a) Estimació de k_{min}.

En primer lloc, sabem que si el mètode iteratiu és convergent, la seva matriu d'iteració B satisfà $\rho(B) < 1$. Si $\bar{x}$ és la solució del sistema lineal $Ax = b$ i $e^{(k)} = x^{(k)} - \bar{x}$, llavors

$$\frac{\|e^{(k)}\|}{\|e^{(0)}\|} \leq \|B^k\|,$$

per a qualsevol norma induïda per una norma vectorial. Si volem que $\|e^{(k)}\| \leq \epsilon\|e^{(0)}\|$, cal que $\|B^k\| \leq \epsilon$ o equivalentment

$$\log(\|B^k\|) \leq \log \epsilon.$$

Si suposem que $(1/k)\log(\|B^k\|) \approx \log(\rho(B))$ (cosa que és certa si k és prou gran), llavors obtenim l'estimació

$$k_{min} \approx \frac{\log \epsilon}{\log(\rho(B))}.$$

Per a obtenir una estimació de $\rho(B)$ podem fer el següent: Definim $\delta^{(k)} = x^{(k+1)} - x^{(k)}$. Si k és gran, llavors $\rho(B) \approx \theta_k$, on $\theta_k = \|\delta^{(k+1)}\|/\|\delta^{(k)}\|$.

b) Criteri de parada per l'error absolut.

Sigui $\|\cdot\|$ una norma matricial induïda per una norma vectorial tal que $\|B\| < 1$. Per la proposició 1.2.1, si volem que $\|e^{(k+1)}\| < \epsilon$, només cal imposar que

$$\frac{\|B\|}{1 - \|B\|}\|\delta^{(k)}\| < \epsilon.$$

També podem estimar el nombre d'iterats necessaris per assolir aquest error usant (1.3):

$$\|e^{(k+1)}\| \leq \frac{\|B\|^{k+1}}{1-\|B\|}\|\delta^{(0)}\|.$$

Fins aquí les estimacions són riguroses. El problema és com estimar $\|B\|$. En el cas de que la matriu sigui diagonal dominant estricte per files i el mètode iteratiu sigui el de Jacobi o Gauss-Seidel, que introduirem a la secció 1.2.4, tenim que $\|B\|_\infty < 1$ i és fàcil obtenir la norma o una fita de la norma de B (vegis la demostració de la proposició 1.2.3). En general, podem donar una estimació (no una fita rigurosa) de la següent manera: Sabem que

$$x^{(k+1)} - x^{(k)} = e^{(k+1)} - e^{(k)} = B(e^{(k)} - e^{(k-1)}) = B(x^{(k)} - x^{(k-1)}).$$

Per tant

$$\frac{\|\delta^{(k)}\|}{\|\delta^{(k-1)}\|} \leq \|B\|.$$

Si suposem que aquesta fita inferior és aproximadament igual a la norma tenim una estimació no rigurosa d'una fita superior $\epsilon^{(k+1)}$ de la norma de $e^{(k+1)}$ usant la (1.2):

$$\epsilon^{(k+1)} = \frac{\|\delta^{(k)}\|^2}{\|\delta^{(k-1)}\| - \|\delta^{(k)}\|}.$$

Tot i que no és realment una fita superior, en molts casos pot ser una bona indicació de l'error real.

c) Criteri de parada per l'error relatiu. Sigui $r^{(k)} = b - Ax^{(k)}$ el residu en el pas k. Si imposem que $\|r^{(k)}\|/\|b\| \leq \epsilon$ llavors obtenim la següent fita de l'error relatiu

$$\frac{\|\bar{x} - x^{(k)}\|}{\|\bar{x}\|} = \frac{\|A^{-1}b - x^{(k)}\|}{\|\bar{x}\|} = \frac{\|A^{-1}r^{(k)}\|}{\|\bar{x}\|} \leq \frac{\|A^{-1}\|\,\|r^{(k)}\|}{\|\bar{x}\|} \leq \kappa(A)\frac{\|r^{(k)}\|}{\|b\|} \leq \epsilon\kappa(A).$$

Veiem que si el nombre de condició és molt gran, l'error relatiu serà molt més gran que la norma del residu.

1.2.3 Una família de mètodes lineals estacionaris

Podem construir mètodes iteratius de la següent manera:

Suposem que el sistema a resoldre és

$$Ax = b \tag{1.4}$$

on $\det A \neq 0$. Aleshores la matriu de coeficients es pot 'trencar', d'infinites maneres, de la forma

$$A = N - P \tag{1.5}$$

on N i P són matrius de la mateixes dimensions que A. Aleshores, el sistema (1.4) s'escriu com

$$Nx = Px + b. \tag{1.6}$$

Començant per un vector arbitrari $x^{(0)}$, definim una successió de vectors $(x_i)_i$, per la recursió

$$Nx^{(k+1)} = Px^{(k)} + b, \qquad i = 1, 2, \ldots. \tag{1.7}$$

Està clar que una de les restriccions que hem d'imposar és

$$\det N \neq 0,$$

ja que llavors el procés iteratiu (1.7) defineix una única successió de vectors per tot $x^{(0)}$ i tot b. Com a qüestió pràctica, N s'haurà d'escollir de tal forma que un sistema de la forma

$$Ny = z$$

pugui èsser resol fàcilment. A més, si volem més precisió, és millor considerar la forma equivalent

$$N(x^{(k+1)} - x^{(k)}) = b - Ax^{(k)}.$$

Observem que si els iterats $(x^{(k)})_k$ convergeixen, i $\bar{x}$ és el seu límit, llavors $A\bar{x} = b$, i per tant $\bar{x}$ és la solució cercada del sistema. Per tant, els mètodes descrits són consistents i la seva matriu d'iteració és $B = N^{-1}P$.

En la propera secció descriurem dos dels mètodes més coneguts d'aquest tipus.

1.2.4 Mètodes de Jacobi i Gauss-Seidel

Presentem aquí dos mètodes molt relacionats. Potser el mètode iteratiu més simple és el **Mètode de Jacobi**:

Suposem que $a_{ii} \neq 0$, per a tot $1 \leq i \leq n$. Cada equació del sistema $Ax = b$ es pot escriure com

$$\sum_{j=1, j\neq i}^{n} a_{ij}x_j + a_{ii}x_i = b_i.$$

Aïllant la variable x_i obtenim

$$x_i = \frac{1}{a_{ii}}\left[b_i - \sum_{j=1, j\neq i}^{n} a_{ij}x_j\right], \qquad i = 1, 2, \ldots, n.$$

Per tant, podem definir l'esquema iteratiu:

$$x_i^{(k+1)} = \frac{1}{a_{ii}}\left[b_i - \sum_{j=1, j\neq i}^{n} a_{ij}x_j^{(k)}\right], \qquad i = 1, 2, \ldots, n. \tag{1.8}$$

S'acostuma a agafar com a condició inicial $x^{(0)} = 0$ (més endavant discutirem això). Notem que trivialment, si la successió convergeix, el seu límit és la solució del sistema.

Una altra manera d'obtenir el mètode de Jacobi és la següent: Diem D a la matriu diagonal de A, L a la matriu triangular inferior estricte de A i U a la matriu triangular superior estricte de A,

és a dir:

$$L = \begin{bmatrix} 0 & 0 & \cdots & 0 \\ a_{21} & 0 & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ a_{n1} & \cdots & a_{n,n-1} & 0 \end{bmatrix},$$

$$D = \begin{bmatrix} a_{11} & 0 & \cdots & 0 \\ 0 & a_{22} & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & a_{n,n} \end{bmatrix},$$

$$U = \begin{bmatrix} 0 & a_{12} & \cdots & a_{1n} \\ 0 & 0 & \ddots & \vdots \\ \vdots & \ddots & \ddots & a_{n-1,n} \\ 0 & \cdots & 0 & 0 \end{bmatrix}.$$

Llavors, $Ax = (L + D + U)x = (L + U)x + Dx = b$ i, equivalentment

$$x = D^{-1}b - D^{-1}(L + U)x.$$

Ovservem que amb la notació de la introducció, $c = D^{-1}b$ i $B = -D^{-1}(L+U)$. A més, el mètode es pot posar com un mètode de la família de la secció anterior posant $N = D$ i $P = -(L+U)$, és a dir

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

Una modificació del Mètode de Jacobi ens proporciona el **mètode de Gauss-Seidel**: Notem que el mètode de Jacobi no usem tota la informació disponible quan calculem l'aproximació de la component i-èssima de l'aproximació de la solució. És a dir, podem escriure el mètode de Jacobi de la següent manera:

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

per observar que els elements $x_j^{(k+1)}$, $j = 1, \ldots, i-1$ ja són coneguts quan calculem $x_i^{(k+1)}$. Si substituim els $x_j^{(k)}$ pels $x_j^{(k+1)}$, $j = 1, \ldots, i-1$ en la fórmula, obtenim el mètode de Gauss-Seidel:

$$x_i^{(k+1)} = \frac{1}{a_{ii}} \left[b_i - \sum_{j=1}^{i-1} a_{ij} x_j^{(k+1)} - \sum_{j=i+1}^{n} a_{ij} x_j^{(k)} \right], \tag{1.9}$$

Si ho escrivim amb notació matricial tenim, com abans que si $Ax = b$ llavors $x = D^{-1}[b - Lx - Ux]$. Si comparem amb la fòrmula anterior, veiem que és equivalent a

$$x^{(k+1)} = D^{-1}[b - Lx^{(k+1)} - Ux^{(k)}].$$

Per tenir-ho en forma standard, cal aïllar $x^{(k+1)}$:

$$Dx^{(k+1)} = b - Lx^{(k+1)} - Ux^{(k)},$$

o

$$(L+D)x^{(k+1)} = b - Ux^{(k)}.$$

Per tant, obtenim

$$x^{(k+1)} = (L+D)^{-1}b - (L+D)^{-1}Ux^{(k)}.$$

També el mètode de Gauss-Seidel forma part de la família de la secció anterior. Aquí $N = D + L$ i $P = -U$.

Nota 1.2.4 *Per implementar els mètodes de Jacobi i Gauss-Seidel usarem* ***sempre*** *els esquemes iteratius (1.8) i (1.9). La deducció de les matrius d'iteració dels mètodes són importants únicament per a determinar la seva convergència.*

Anem a discutir la convergència d'aquests mètodes.

Proposició 1.2.2 *Si A és diagonal dominant per files en sentit estricte, llavors el mètode de Jacobi és convergent.*

Demostracio:
Una matriu $A = (a_{ij})_{1\leq i,j\leq n}$ és **diagonal dominant per files en sentit estricte** quan

$$|a_{ii}| > \sum_{j=1, j\neq i}^{n} |a_{ij}|, \qquad i = 1, \ldots, n.$$

Per tant

$$\max_{1\leq i\leq n} \frac{1}{|a_{ii}|} \sum_{j=1, j\neq i}^{n} |a_{ij}| = \alpha < 1.$$

Recordem que pel mètode de Jacobi la matriu d'iteració és $B = -D^{-1}(L+U)$. Llavors

$$\|B\|_\infty = \max_{1\leq i\leq n} \sum_{j=1, j\neq i}^{n} \frac{|a_{ij}|}{|a_{ii}|} = \alpha < 1.$$

Aplicant el Teorema 1.2.1, tenim demostrada la proposició. □

Nota 1.2.5 *Es pot demostrar que si una matriu és diagonal dominant llavors es pot aplicar el mètode de Jacobi (trivial!) i que és regular, demostrant que el nucli consisteix en el vector zero.*

Nota 1.2.6 *En la demostració es calcula $\|B\|_\infty$ i és demostra que és menor que 1. Per tant, podem donar un criteri de parada rigoròs per l'error absolut si agafem la norma del màxim.*

Nota 1.2.7 *Si la matriu és diagonal dominant en sentit estricte per columnes, llavors el mètode de Jacobi també és convergent.*

Proposició 1.2.3 *Si A és diagonal dominant per files en sentit estricte, llavors el mètode de Gauss-Seidel és convergent.*

Demostració:
Recordem que, en aquest cas, $B = -(L+D)^{-1}U$. Sabem que

$$\|B\|_\infty = \max_{\|x\|_\infty=1} \|Bx\|_\infty.$$

Sigui x tal que $\|x\|_\infty = 1$. Aleshores si $y = Bx$, prenem k tal que $\|y\|_\infty = |y_k|$. Tenim que $(L+D)y = -Ux$ o

$$Dy = -Ly - Ux.$$

L'equació k-èssima corresponent és

$$a_{kk}y_k = -\sum_{j=1}^{k-1} a_{kj}y_j - \sum_{j=k+1}^{n} a_{kj}x_j.$$

Per tant,

$$\|y\|_\infty = |y_k| \leq \sum_{j=1}^{k-1} \frac{|a_{kj}|}{|a_{kk}|}\|y\|_\infty + \sum_{j=k+1}^{n} \frac{|a_{kj}|}{|a_{kk}|}\|x\|_\infty.$$

Si anomenem

$$s_k = \sum_{j=1}^{k-1} \frac{|a_{kj}|}{|a_{kk}|}, \qquad r_k = \sum_{j=k+1}^{n} \frac{|a_{kj}|}{|a_{kk}|},$$

aleshores

$$\|y\|_\infty \leq s_k\|y\|_\infty + r_k.$$

Per tant,

$$\|y\|_\infty \leq \frac{r_k}{1-s_k},$$

ja que $0 < s_k < 1$ per ser diagonal dominant estricte. Així

$$\|B\|_\infty \leq \max_{1\leq k\leq n} \frac{r_k}{1-s_k} < 1,$$

degut a que $r_k + s_k < 1$, per a tot k, per ser diagonal dominant estricte. Finalment apliquem novament el Teorema 1.2.1. □

Nota 1.2.8 *En la demostració es troba una fita superior menor que 1 de $\|B\|_\infty$. Per tant, podem donar un criteri de parada rigoròs per l'error absolut si agafem la norma del màxim.*

Exemple 1.2.1 *Volem resoldre el sistema:*

$$\begin{bmatrix} 7 & -6 \\ -8 & 9 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 3 \\ -4 \end{bmatrix}.$$

En primer lloc, observem que la solució és $(x_1, x_2) = (1/5, -4/15)$.

Si usem el mètode de Jacobi tenim

$$\begin{aligned} x_1^{(k+1)} &= \frac{6}{7}x_2^{(k)} + \frac{3}{7}, \\ x_2^{(k+1)} &= \frac{8}{9}x_1^{(k)} - \frac{4}{9}. \end{aligned}$$

Per tant, la matriu B en aquest cas és

$$B = \begin{bmatrix} 0 & \frac{6}{7} \\ \frac{8}{9} & 0 \end{bmatrix},$$

i $\|B\|_\infty = \frac{8}{9} \approx 0.889$, $\rho(B) = \sqrt{16/21} \approx 0.87287$. Si usem el mètode de Gauss-Seidel tenim:

$$\begin{aligned} x_1^{(k+1)} &= \frac{6}{7}x_2^{(k)} + \frac{3}{7}, \\ x_2^{(k+1)} &= \frac{8}{9}x_1^{(k+1)} - \frac{4}{9}, \end{aligned}$$

o, substituint en la segona equació el valor de $x_1^{(k+1)}$ en la primera,

$$\begin{aligned} x_1^{(k+1)} &= \frac{6}{7}x_2^{(k)} + \frac{3}{7}, \\ x_2^{(k+1)} &= \frac{16}{21}x_2^{(k)} - \frac{4}{63}. \end{aligned}$$

En aquest cas, la matriu B és

$$B = \begin{bmatrix} 0 & \frac{6}{7} \\ 0 & \frac{16}{21} \end{bmatrix},$$

i per tant, $\|B\|_\infty = \frac{6}{7} \approx 0.857$, $\rho(B) = 16/21 \approx 0.76190$. Això vol dir que en aquest cas, el mètode de Gauss-Seidel convergeix més rapidament a la solució que el de Jacobi. En qualsevol cas, la convergència és molt lenta. Després de 50 iterats, l'error en el mètode de Jacobi en la norma del suprem és aproximadament $3.3 \cdot 10^{-4}$ i en el de Gauss-Seidel $5 \cdot 10^{-7}$. En les properes seccions veurem mètodes que acceleren la convergència.

Nota 1.2.9 *En general, no podem dir que el mètode de Gauss-Seidel convergeixi més ràpidament que el de Jacobi, ni que la convergència d'un mètode impliqui la de l'altre*[1].

1.2.5 Mètode de sobrerelaxació (SOR)

En general un mètode de (sobre)relaxació consisteix en una modificació d'un mètode iteratiu amb el que es pretèn millorar la velocitat de convergència. Nosaltres considerarem el mètode de Gauss-Seidel, encara que també és possible construir un mètode associat al mètode de Jacobi.

Recordem que el mètode de Gauss-Seidel el podem escriure com

$$x_i^{(k+1)} = \frac{1}{a_{ii}} \left[b_i - \sum_{j<i} a_{ij} x_j^{(k+1)} - \sum_{j>i} a_{ij} x_j^{(k)} \right], \qquad i = 1, \ldots, n.$$

Aquest mètode també el podem escriure com $x_i^{(k+1)} = x_i^{(k)} + r_i^{(k)}$, on

$$r_i^{(k)} = \frac{1}{a_{ii}} \left[b_i - \sum_{j<i} a_{ij} x_j^{(k+1)} - \sum_{j\geq i} a_{ij} x_j^{(k)} \right], \qquad i = 1, \ldots, n.$$

[1] Veure A. Quarteroni et al: Numerical Mathematics, exemple 4.2, pàgina 132

Per a modificar el mètode, prenem un valor $\omega \in \mathbb{R}$ (anomenat **factor de relaxació**) i definim, donat un valor inicial $x^{(0)}$,

$$x_i^{(k+1)} = x_i^{(k)} + \omega r_i^{(k)}, \qquad i = 1, \ldots, n,$$

Aquest mètode l'anomenem **mètode de sobrerelaxació** (successive over-relaxation – SOR). Si $\omega = 1$ recuperem el mètode de Gauss-Seidel. Per a calcular la matriu d'iteració del mètode, l'escrivim en forma matricial: el mètode de Gauss-Seidel es pot escriure:

$$x^{(k+1)} = D^{-1}(b - Lx^{(k+1)} - Ux^{(k)}) = x^{(k)} + D^{-1}[b - Lx^{(k+1)} - (D+U)x^{(k)}],$$

i per tant el mètode SOR s'escriurà:

$$x^{(k+1)} = x^{(k)} + \omega D^{-1}[b - Lx^{(k+1)} - (D+U)x^{(k)}].$$

Aïllant $x^{(k+1)}$, obtenim:

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

Per tant,

$$x^{(k+1)} = B_{SOR}(\omega)x^{(k)} + c(\omega),$$

on

$$B_{SOR}(\omega) = (I + \omega D^{-1}L)^{-1}[(1-\omega)I - \omega D^{-1}U], \qquad c(\omega) = \omega(I + \omega D^{-1}L)^{-1}D^{-1}b.$$

Nota 1.2.10 *Per a obtenir aquestes fórmules hem usat la descomposició $A = L + D + U$. Una altra opció és $A = D[D^{-1}L + I + D^{-1}U) = D(\bar{L} + I + \bar{U})$. Aleshores SOR s'escriu:*

$$B_{SOR}(\omega) = (I + \omega\bar{L})^{-1}[(1-\omega)I - \omega\bar{U}].$$

Amb aquesta notació, quan $\omega = 1$ obtenim el mètode de Gauss-Seidel:

$$B_{GS} = -(I + \bar{L})^{-1}\bar{U}.$$

Una condició necesària per a tenir convergència dels mètodes SOR la dona el següent teorema:

Teorema 1.2.3 *El radi espectral de la matriu de SOR satisfà $\rho(B_{SOR}(\omega)) \geq |\omega - 1|$, i per tant, si $\omega \in \mathbb{R}$, l mètode només pot convergir per $0 < \omega < 2$.*

Demostraciò:

Com que

$$\det(B_{SOR}) = \det(I + \omega\bar{L})^{-1}\det((1-\omega)I - \omega\bar{U}) = (1-\omega)^n,$$

i $\det(B_{SOR}) = \lambda_1 \cdots \lambda_n$, on λ_i, $i = 1, \ldots, n$, són els valors propis de B_{SOR}, aleshores

$$|1-\omega|^n = |\lambda_1| \cdots |\lambda_n| \leq \rho(B_{SOR})^n,$$

el que prova el que voliem. □

Hi ha casos especials en els que es poden obtenir condicions suficients de convergència dels mètodes SOR:

Teorema 1.2.4 *(Ostrowski-Reich) Si A és simètrica i definida positiva, llavors SOR convergeix per a tot $\omega \in (0,2)$.*

Corol·lari 1.2.1 *Si A és simètrica i definida positiva, el mètode de Gauss-Seidel és convergent.*

En general, no hi ha una fórmula general per trobar el valor òptim del paràmetre de relaxació ω inclús en el cas en el que A és simètrica i definida positiva. Tot i així, en determinats casos tenim més informació:

Teorema 1.2.5 *Sigui A una matriu real, simètrica, definida positiva i tridiagonal a blocs:*

$$A = \begin{pmatrix} D_1 & U_1 & & & 0 \\ L_2 & D_2 & U_2 & & \\ & L_3 & D_3 & U_3 & \\ & \ddots & \ddots & \ddots & \\ & & L_{n-1} & D_{n-1} & U_{n-1} \\ 0 & & & L_n & D_n \end{pmatrix},$$

on D_i són matrius diagonals. Llavors, $\rho(B_{GS}) = \rho(B_J)^2$ i el valor òptim del paràmetre de relaxació és

$$\tilde{\omega} = \frac{2}{1 + (1 - \rho(B_{GS}))^{1/2}}, \qquad \rho(B_{GS}) < 1,$$

on $\rho(B_J)$ és el radi espectral de la matriu d'iteració del mètode de Jacobi, corresponent a A, i B_{GS} el del mètode de Gauss-Seidel. El valor òptim de $\rho(B_\omega)$ és $\rho(B_{\tilde{\omega}}) = \tilde{\omega} - 1$.

En general, no es pot donar una fórmula general per a aproximar $\tilde{\omega}$. Si cal resoldre molts sistemes amb la mateixa matriu, pot pagar la pena invertir un cert esforç per aproximar $\tilde{\omega}$.

Nota 1.2.11 *Els mètodes iteratius poden ser molt adequats per a sistemes de dimensió gran amb una certa estructura i amb pocs elements diferents de zero (matrius escasses). A més a més, si es coneix una aproximació a la solució, i no hi ha un requeriment molt gran de precisió, els mètodes iteratius són altament recomenables.*

Nota 1.2.12 *Una de les dificultats per aplicar els mètodes iteratius és conèixer la convergència a priori. Tot i així, els casos en el que tenim informació sobre la convergència són molt rellevants, perquè apareixen de manera natural en la resolució númerica d'equacions en derivades parcials (veure l'exemple de la introducció).*

Nota 1.2.13 *(implementació de mètodes iteraius) Notem que els mètodes iteratius que hem vist, a diferència dels directes, preserven el caràcter escas de la matriu. Quan s'implementen aquests mètodes per a matrius escasses, especialment quan la dimensió de la matriu és gran, només es necessita una funció que calculi el producte d'una matriu adient fixada per un vector. En cap cas s'ha de reservar memòria per la matriu.*

1.2.6 Mètodes de Minimització

El problema dels mètodes SOR i d'altres de similars, és que pot ser complicada l'elecció del paràmetre ω. Però si la matriu A és real, simètrica i definida positiva, el paràmetre òptim d'acceleració (o relaxació) es pot calcular a cada pas (donant lloc a un mètode no estacionari en general). En primer lloc observem que el problema de resoldre $Ax = b$ és equivalent a minimitzar la funció quadràtica:

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

En efecte, sabem que un mínim de Q ha de ser un punt crític de Q, és a dir, un zero de

$$DQ(x) = x^T A - b^T,$$

on el seu Hessià és $H(x) = A$, (o alternativament $D^2Q(x)(h_1, h_2) = h_1^T Ah_2$). Com que A és definida positiva, veiem que la solució de $Ax = b$ és el mínim de $Q(x)$.

Hi ha molts mètodes iteratius per a minimitzar la funció Q. Molts dells són de la forma:

$$x^{(k+1)} = x^{(k)} - \alpha_k p^{(k)}, \qquad k = 0, 1, \ldots,$$

on els vectors $p^{(k)} \neq 0$ determinen direccions i els escalars α_k la distància a moure's en la direcció de $p^{(k)}$. Hi ha una gran varietat de mètodes per seleccionar els α_k i els p_k. Potser la manera més natural d'escollir els α_k és minimitzar Q en la direcció de $p^{(k)}$. Aleshores

$$Q(x^{(k)} - \alpha_k p^{(k)}) = \min_{\alpha} Q(x^{(k)} - \alpha p^{(k)}).$$

Si fixem $x^{(k)}$ i $p^{(k)}$, això és un problema de minimització unidimensional, que es pot resoldre de forma explícita. Si anomenem $x = x^{(k)}$ i $p = p^{(k)}$ tenim que

$$q(\alpha) = Q(x-\alpha p) = \frac{1}{2}(x-\alpha p)^T A(x-\alpha p) - b^T(x-\alpha p) = \frac{1}{2}x^T Ax - \alpha p^T Ax + \frac{1}{2}\alpha^2 p^T Ap + \alpha p^T b - b^T x =$$

$$\frac{1}{2}p^T Ap\alpha^2 - p^T(Ax - b)\alpha + \frac{1}{2}x^T(Ax - 2b).$$

Com que A és definida positiva, tenim que $p^T Ap > 0$ i el mínim s'assoleix quan $q'(\alpha) = 0$, és a dir:

$$\alpha_k = \frac{(p^{(k)})^T(Ax^{(k)} - b)}{(p^{(k)})^T Ap^{(k)}}. \tag{1.10}$$

D'aquesta manera podem definir el mètode iteratiu no estacionari següent: Donats $x^{(0)}$, $p^{(k)}$, $k \geq 0$:

$$x^{(k+1)} = x^{(k)} - \alpha_k p^{(k)}, \quad \alpha_k = \frac{(p^{(k)})^T(Ax^{(k)} - b)}{(p^{(k)})^T Ap^{(k)}}, \qquad k = 0, 1, 2, \ldots,$$

que anomenem **mètode de minimització**.

Relaxació univariant

Si escollim adequadament les direccions $p^{(k)}$, podem obtenir un vincle entre els mètodes de minimització i el mètode de Gauss-Seidel.

Sigui e_i el vector i-èssim de la base canònica. Una de les eleccions més simples de les direccions és agafar els vectors de la base canònica $e_1, \ldots, e_n$ de manera cíclica:

$$p^{(0)} = e_1, p^{(1)} = e_2, \ldots, p^{(n-1)} = e_n, p^{(n)} = e_1, \ldots$$

Notem que $e_i^T A e_i = a_{ii}$, i que $e_i^T (Ax - b) = \sum_{j=1}^n a_{ij} x_j - b_i$. Per tant, si $p^{(k)} = e_i$ i α_k s'agafa com abans, el següent iterat és

$$x^{(k+1)} = x^{(k)} - \alpha_k e_i = x^{(k)} - \frac{1}{a_{ii}} \left[\sum_{j=1}^n a_{ij} x_j^{(k)} - b_i \right] e_i.$$

Notem que $x^{(k+1)}$ i $x^{(k)}$ només difereixen en la component i. De fet, aquest esquema equival a minimitzar Q en la variable x_i mentre les altres variables mantenen el seus valors.

Considerem ara els primers n passos, que escriurem només per les components que varien, i notem que $x_j^{(k)} = x_j^{(0)}$, fins que es canvia la component j :

$$x_i^{(i)} = x_i^{(0)} - \alpha_i = x_i^{(0)} - \frac{1}{a_{ii}} \left(\sum_{j=1}^{i-1} a_{ij} x_j^{(j)} + \sum_{j=i}^n a_{ij} x_j^{(0)} - b_i \right) = \frac{1}{a_{ii}} \left(b_i - \sum_{j=1}^{i-1} a_{ij} x_j^{(j)} - \sum_{j=i+1}^n a_{ij} x_j^{(0)} \right),$$

$$i = 1, \ldots, n.$$

Observem que els primers n passos de la relaxació univariant constitueixen una iteració de Gauss-Seidel. Podem dir que el mètode de Gauss-Seidel és equivalent a n passos consecutius de la relaxació univariant.

Sobrerelaxació

Per a qualsevol mètode de la forma

$$x^{(k+1)} = x^{(k)} - \alpha_k p^{(k)},$$

podem afegir un paràmetre de relaxació ω a la minimització, de tal forma que α_k s'agafa com

$$\alpha_k = \omega \tilde{\alpha}_k,$$

on $\tilde{\alpha}_k$ és ara el valor de α que minimitza Q en la direcció $p^{(k)}$. És fàcil veure que $Q(x^{(k)} - \omega \tilde{\alpha}_k p^{(k)}) < Q(x^{(k)})$, si $\omega \in (0, 2)$, i que $Q(x^{(k)} - \omega \tilde{\alpha}_k p^{(k)}) \geq Q(x^{(k)})$, altrament.

Mètode del descens més ràpid o mètode del gradient

Per a una funció g de n variables, $-\nabla g(x)$ dona la direcció de màxim decreixement local de la funció g a x, Per tant, un elecció natural del vector $p^{(k)}$ d'un algorisme de minimització és

$$p^{(k)} = \nabla Q(x^{(k)}) = Ax^{(k)} - b,$$

i aquesta defineix el **Mètode de descens més ràpid**, també conegut com a **Mètode de Richardson o del gradient**. Notem que si normalitzem A per a que tingui elements de la diagonal iguals a 1 i $\alpha_k = 1$, aleshores el mètode es redueix a l'iteració de Jacobi (exercici).

Si definim $r^{(k)} = b - Ax^{(k)}$, podem descriure aquest mètode de la següent manera: Donat $x^{(0)} \in \mathbb{R}^n$, definim $r^{(0)} = b - Ax^{(0)}$, i, per $k = 0, 1, \ldots$ calculem

$$\begin{aligned} \alpha_k &= \frac{(r^{(k)})^T r^{(k)}}{(r^{(k)})^T A r^{(k)}}, \\ x^{(k+1)} &= x^{(k)} + \alpha_k r^{(k)}, \\ r^{(k+1)} &= r^{(k)} - \alpha_k A r^{(k)}. \end{aligned}$$

Tot i que sembla la millor elecció, aquest mètode convergeix molt lentament en general. En efecte, tenim el següent teorema[2]

Teorema 1.2.6 *Sigui A una matriu simètrica i definida positiva; aleshores el mètode del gradient és convergent per a qualsevol elecció de la dada inicial $x^{(0)}$. A més,*

$$\|x^{(k+1)} - \bar{x}\|_A \leq \frac{\kappa_2(A) - 1}{\kappa_2(A) + 1} \|x^{(k)} - \bar{x}\|_A,$$

on $\bar{x}$ és la solució del sistema $Ax = b$, $\kappa_2(A)$ és el nombre de condició de A amb la norma euclidiana i $\|x\|_A = \sqrt{x^T A x}$.

Aquest teorema ens diu que si el nombre de condició de A és gran, el mètode convergirà molt lentament.

Mètodes de direccions conjugades

Una classe molt important de mètodes, anomenats **Mètodes de direccions conjugades**, surgeixen quan hi ha n vectors $p^{(0)}, \ldots, p^{(n-1)}$ que satisfàn

$$(p^{(i)})^T A p^{(j)} = 0, \qquad i \neq j. \tag{1.11}$$

Tals vectors són ortogonals respecte del producte escalar

$$\langle x, y \rangle = x^T A y,$$

definit per A, i s'anomenen **conjugats** respecte de A. Una de les propietats bàsiques dels mètodes de direccions conjugades és:

Teorema 1.2.7 *(direccions conjugades) Si A és una matriu real, simètrica i definida positiva $n \times n$ i $p^{(0)}, \ldots, p^{(n-1)}$ són vectors no nuls que satisfan (1.11), aleshores per a qualsevol $x^{(0)}$, els iterats $x^{(k+1)} = x^{(k)} - \alpha_k p^{(k)}$, on α_k està definit com a (1.10), convergeixen a la solució exacta de $Ax = b$ en no més de n passos.*

[2]Veure A. Quarteroni et al: Numerical Mathematics, teorema 4.10. pàgina 149.

Demostració:
Notem que

$$(Ax^{(k+1)} - b)^T p^{(j)} = (Ax^{(k)} - \alpha_k A p^{(k)} - b)^T p^{(j)} = (Ax^{(k)} - b)^T p^{(j)} - \alpha_k (Ap^{(k)})^T p^{(j)}.$$

Usant l'A-ortogonalitat dels $p^{(k)}$ i la definició de α_k, obtenim

$$(Ax^{(k+1)} - b)^T p^{(j)} = \begin{cases} (Ax^{(k)} - b)^T p^{(j)} & \text{si} \quad j < k, \\ 0 & \text{si} \quad j = k. \end{cases}$$

Per tant,

$$(Ax^{(n)} - b)^T p^{(j)} = (Ax^{(n-1)} - b)^T p^{(j)} = \cdots = (Ax^{(j+1)} - b)^T p^{(j)} = 0,$$

per $j = 0, \ldots, n-1$. Com que $p^{(0)}, \ldots, p^{(n-1)}$ són linealment independents, necessariament $Ax^{(n)} - b = 0$. Podria passar que $Ax^{(m)} - b = 0$, per a $m < n$, amb el que també tindriem la solució. □

Tot i que el mètode del gradient conjugat és un mètode directe (s'obté la solució en un nombre finit de passos), a la pràctica es pot usar com un mètode iteratiu, i convergeix més ràpidament que el mètode del gradient[3]

Teorema 1.2.8 *Sigui A una matriu $n \times n$ simètrica i definida positiva. El mètode del gradient conjugat per resoldre $Ax = b$ satisfà per $k < n$:*

$$\|x^{(k)} - \bar{x}\|_A \leq \frac{2c^k}{1 + c^{2k}} \|x^{(0)} - \bar{x}\|_A, \qquad c = \frac{\sqrt{\kappa_2(A)} - 1}{\sqrt{\kappa_2(A)} + 1}.$$

[3]veure A. Quarteroni et al, Numerical Methods, teorema 4.12, pàgina 155.