1.5 Méthodes itératives

Les méthodes directes sont très efficaces : elles donnent la solution exacte (aux erreurs d'arrondi près) du système linéaire considéré. Elles ont l'inconvénient de nécessiter une assez grande place mémoire car elles nécessitent le stockage de toute la matrice en mémoire vive. Si la matrice est pleine, c.à.d. si la plupart des coefficients de la matrice sont non nuls et qu'elle est trop grosse pour la mémoire vive de l'ordinateur dont on dispose, il ne reste plus qu'à gérer habilement le "swapping" c'est–à–dire l'échange de données entre mémoire disque et mémoire vive pour pouvoir résoudre le système.
Cependant, si le système a été obtenu à partir de la discrétisation d'équations aux dérivés partielles, il est en général "creux", c.à. d. qu'un grand nombre des coefficients de la matrice du système sont nuls ; de plus la matrice a souvent une structure "bande", *i.e.* les éléments non nuls de la matrice sont localisés sur certaines diagonales. On a vu au chapitre précédent que dans ce cas, la méthode de Choleski "conserve le profil" (voir à ce propos page 45). Si on utilise une méthode directe genre Choleski, on aura donc besoin de la place mémoire pour stocker la strucutre bande.
Lorsqu'on a affaire à de très gros systèmes issus par exemple de l'ingénierie (calcul des structures, mécanique des fluides, ...), où n peut être de l'ordre de plusieurs milliers, on cherche à utiliser des méthodes nécessitant le moins de mémoire possible. On a intérêt dans ce cas à utiliser des méthodes itératives. Ces méthodes ne font appel qu'à des produits matrice vecteur, et ne nécessitent donc pas le stockage du profil de la matrice mais uniquement des termes non nuls. Par exemple, si on a seulement 5 diagonales non nulles dans la matrice du système à résoudre, système de n équations et n inconnues, la place mémoire nécessaire pour un produit matrice vecteur est $6n$. Ainsi pour les gros systèmes, il est souvent avantageux d'utiliser des méthodes itératives qui ne donnent pas toujours la solution exacte du système en un nombre fini d'itérations, mais qui donnent une solution approchée à coût moindre qu'une méthode directe, car elles ne font appel qu'à des produits matrice vecteur.

Remarque 1.45 (Sur la méthode du gradient conjugué)**.**
Il existe une méthode itérative "miraculeuse" de résolution des systèmes linéaires lorsque la matrice A est symétrique définie positive : c'est la méthode du gradient conjugué. Elle est miraculeuse en ce sens qu'elle donne la solution exacte du système $Ax = b$ en un nombre fini d'opérations (en ce sens c'est une méthode directe) : moins de n itérations où n est l'ordre de la matrice A, bien qu'elle ne nécessite que des produits matrice vecteur ou des produits scalaires. La méthode du gradient conjugué est en fait une méthode d'optimisation pour la recherche du minimum dans $\mathbb{R}^n$ de la fonction de $\mathbb{R}^n$ dans $\mathbb{R}$ définie par : $f(x) = \frac{1}{2}Ax \cdot x - b \cdot x$. Or on peut montrer que lorsque A est symétrique définie positive, la recherche de x minimisant f dans $\mathbb{R}^n$ est équivalent à la résolution du système $Ax = b$. (Voir paragraphe 3.2.2 page 213.) En fait, la méthode du gradient conjugué n'est pas si miraculeuse que cela en pratique : en effet, le nombre n est en général très grand et on ne peut en géneral pas envisager d'effectuer un tel nombre d'itérations pour résoudre le système. De plus, si on utilise la méthode du gradient conjugué brutalement, non seulement elle ne donne pas la solution en n itérations en raison de l'accumulation des erreurs d'arrondi, mais plus la taille du système croît et plus le nombre d'itérations nécessaires devient élevé. On a alors recours aux techniques de "préconditionnement". Nous reviendrons sur ce point au chapitre 3.
La méthode itérative du gradient à pas fixe, qui est elle aussi obtenue comme méthode de minimisation de la fonction f ci-dessus, fait l'objet de l'exercice 53 page 109 et du théorème 3.19 page 222.

1.5.1 Définition et propriétés

Soit $A \in \mathcal{M}_n(\mathbb{R})$ une matrice inversible et $\boldsymbol{b} \in \mathbb{R}^n$, on cherche toujours ici à résoudre le système linéaire (1.1) c'est–à–dire à trouver $\boldsymbol{x} \in \mathbb{R}^n$ tel que $A\boldsymbol{x} = \boldsymbol{b}$, mais de façon itérative, c.à.d. par la construction d'une suite.

Définition 1.46 (Méthode itérative)**.** *On appelle méthode itérative de résolution du système linéaire* (1.1) *une méthode qui construit une suite $(\boldsymbol{x}^{(k)})_{k\in\mathbb{N}}$ (où l'itéré $\boldsymbol{x}^{(k)}$ est calculé à partir des itérés $\boldsymbol{x}^{(0)} \dots \boldsymbol{x}^{(k-1)}$ censée converger vers x solution de* (1.1)*).*

Bien sûr, on souhaite que cette suite converge vers la solution $\boldsymbol{x}$ du système.

Définition 1.47 (Méthode itérative convergente)**.** *On dit qu'une méthode itérative est convergente si pour tout choix initial $\boldsymbol{x}^{(0)} \in \mathbb{R}^n$, on a :*

$$\boldsymbol{x}^{(k)} \longrightarrow \boldsymbol{x} \text{ quand } k \to +\infty$$

Enfin, on veut que cette suite soit simple à calculer. Une idée naturelle est de travailler avec une matrice P inversible qui soit "proche" de A, mais plus facile que A à inverser. On appelle matrice de préconditionnement cette matrice P. On écrit alors $A = P - (P - A) = P - N$ (avec $N = P - A$), et on réécrit le système linéaire $A\boldsymbol{x} = \boldsymbol{b}$ sous la forme

$$P\boldsymbol{x} = (P - A)\boldsymbol{x} + \boldsymbol{b} = N\boldsymbol{x} + \boldsymbol{b}. \tag{1.90}$$

Cette forme suggère la construction de la suite $(\boldsymbol{x}^{(k)})_{k\in\mathbb{N}}$ à partir d'un choix initial $\boldsymbol{x}^{(0)}$ donné, par la formule suivante :

$$\begin{aligned} P\boldsymbol{x}^{(k+1)} &= (P - A)\boldsymbol{x}^{(k)} + \boldsymbol{b} \\ &= N\boldsymbol{x}^{(k)} + \boldsymbol{b}, \end{aligned} \tag{1.91}$$

ce qui peut également s'écrire :.

$$\boldsymbol{x}^{(k+1)} = B\boldsymbol{x}^{(k)} + \boldsymbol{c}, \text{ avec } B = P^{-1}(P - A) = \mathrm{Id} - P^{-1}A = P^{-1}N \text{ et } \boldsymbol{c} = P^{-1}\boldsymbol{b}. \tag{1.92}$$

Remarque 1.48 (Convergence vers $A^{-1}\boldsymbol{b}$)**.** *Si $P\boldsymbol{x}^{(k+1)} = (P - A)\boldsymbol{x}^{(k)} + b$ pour tout $k \in \mathbb{N}$ et $\boldsymbol{x}^{(k)} \longrightarrow \bar{\boldsymbol{x}}$ quand $k \longrightarrow +\infty$ alors $P\bar{\boldsymbol{x}} = (P - A)\bar{\boldsymbol{x}} + b$, et donc $A\bar{\boldsymbol{x}} = \boldsymbol{b}$, c.à.d. $\bar{\boldsymbol{x}} = \boldsymbol{x}$. En conclusion, si la suite converge, alors elle converge bien vers la solution du système linéaire.*

On introduit l'erreur d'approximation $\boldsymbol{e}^{(k)}$ à l'itération k, définie par

$$\boldsymbol{e}^{(k)} = \boldsymbol{x}^{(k)} - \boldsymbol{x}, \ k \in \mathbb{N} \tag{1.93}$$

où $\boldsymbol{x}^{(k)}$ est construit par (1.92) et $\boldsymbol{x} = A^{-1}\boldsymbol{b}$. Il est facile de vérifier que $\boldsymbol{x}^{(k)} \to \boldsymbol{x} = A^{-1}\boldsymbol{b}$ lorsque $k \to +\infty$ si et seulement si $\boldsymbol{e}^{(k)} \to \boldsymbol{0}$ lorsque $k \to +\infty$

Lemme 1.49. *La suite $(\boldsymbol{e}^{(k)})_{k\in\mathbb{N}}$ définie par* (1.93) *est également définie par*

$$\begin{aligned} \boldsymbol{e}^{(0)} &= \boldsymbol{x}^{(0)} - \boldsymbol{x} \\ \boldsymbol{e}^{(k)} &= B^k\boldsymbol{e}^{(0)} \end{aligned} \tag{1.94}$$

DÉMONSTRATION – Comme $\boldsymbol{c} = P^{-1}\boldsymbol{b} = P^{-1}A\boldsymbol{x}$, on a

$$\boldsymbol{e}^{(k+1)} = \boldsymbol{x}^{(k+1)} - \boldsymbol{x} = B\boldsymbol{x}^{(k)} - \boldsymbol{x} + P^{-1}A\boldsymbol{x} \tag{1.95}$$

$$= B(\boldsymbol{x}^{(k)} - \boldsymbol{x}). \tag{1.96}$$

Par récurrence sur k,

$$\boldsymbol{e}^{(k)} = B^k(\boldsymbol{x}^{(0)} - \boldsymbol{x}), \quad \forall k \in \mathbb{N}. \tag{1.97}$$

■

Théorème 1.50 (Convergence de la suite)**.** *Soit A et $P \in \mathcal{M}_n(\mathbb{R})$ des matrices inversibles. Soit $\boldsymbol{x}^{(0)}$ donné et soit $(\boldsymbol{x}^{(k)})_{k\in\mathbb{N}}$ la suite définie par* (1.92)*.*

1. *La suite $(\boldsymbol{x}^{(k)})_{k\in\mathbb{N}}$ converge, quel que soit $\boldsymbol{x}^{(0)}$, vers $\boldsymbol{x} = A^{-1}\boldsymbol{b}$ si et seulement si $\rho(B) < 1$.*
2. *La suite $(\boldsymbol{x}^{(k)})_{k\in\mathbb{N}}$ converge, quel que soit $\boldsymbol{x}^{(0)}$, si et seulement si il existe une norme induite notée $\|\cdot\|$ telle que $\|B\| < 1$.*

DÉMONSTRATION –

1. On a vu que la suite $(\boldsymbol{x})^{(k)})_{k\in\mathbb{N}}$ définie par (1.92) converge vers $\boldsymbol{x} = A^{-1}\boldsymbol{b}$ si et seulement si la suite $\boldsymbol{e}^{(k)}$ définie par (1.94) tend vers $\mathbf{0}$. On en déduit par le lemme 1.33 que la suite $(\boldsymbol{x})^{(k)})_{k\in\mathbb{N}}$ converge (vers $\boldsymbol{x}$), pour tout $x^{(0)}$, si et seulement si $\rho(B) < 1$.
2. Si il existe une norme induite notée $\|\cdot\|$ telle que $\|B\| < 1$, alors en vertu du corollaire 1.33, $\rho(B) < 1$ et donc la méthode converge pour tout $x^{(0)}$.
   Réciproquement, si la méthode converge alors $\rho(B) < 1$, et donc il existe $\eta > 0$ tel que $\rho(B) = 1 - \eta$. Prenons maintenant $\varepsilon = \frac{\eta}{2}$ et appliquons la proposition 1.32 : il existe une norme induite $\|\cdot\|$ telle que $\|B\| \leq \rho(B) + \varepsilon < 1$, ce qui démontre le résultat.

■

Pour trouver des méthodes itératives de résolution du système (1.1), on cherche donc une décomposition de la matrice A de la forme : $A = P - (P - A) = P - N$, où P est inversible et telle que le système $Py = d$ soit un système facile à résoudre (par exemple P diagonale ou triangulaire).

Estimation de la vitesse de convergence Soit $\boldsymbol{x}^{(0)} \in \mathbb{R}^n$ donné et soit $(\boldsymbol{x}^{(k)})_{k\in\mathbb{N}}$ la suite définie par (1.92). On a vu que, si $\rho(B) < 1$, $\boldsymbol{x}^{(k)} \to x$ quand $k \to \infty$, où $\boldsymbol{x}$ est la solution du système $A\boldsymbol{x} = b$. On montre à l'exercice 71 page 138 que (sauf cas particuliers)

$$\frac{\|\boldsymbol{x}^{(k+1)} - \boldsymbol{x}\|}{\|\boldsymbol{x}^{(k)} - \boldsymbol{x}\|} \longrightarrow \rho(B) \quad \text{lorsque } k \to +\infty,$$

indépendamment de la norme choisie sur $\mathbb{R}^n$. Le rayon spectral $\rho(B)$ de la matrice B est donc une bonne estimation de la vitesse de convergence. Pour estimer cette vitesse de convergence lorsqu'on ne connaît pas $\boldsymbol{x}$, on peut utiliser le fait (voir encore l'exercice 71 page 138) qu'on a aussi

$$\frac{\|\boldsymbol{x}^{(k+1)} - \boldsymbol{x}^{(k)}\|}{\|\boldsymbol{x}^{(k)} - \boldsymbol{x}^{(k-1)}\|} \longrightarrow \rho(B) \ : \ \text{lorsque } k \to +\infty,$$

ce qui permet d'évaluer la vitesse de convergence de la méthode par le calcul des itérés courants.

1.5.2 Quelques exemples de méthodes itératives

Une méthode simpliste

Le choix le plus simple pour le système $Px = (P - A)x + b$ soit facile à résoudre (on rappelle que c'est un objectif dans la construction d'une méthode itérative) est de prendre pour P la matrice identité (qui est très facile à inverser !). Voyons ce que cela donne sur la matrice

$$A = \begin{bmatrix} 2 & -1 \\ -1 & 2 \end{bmatrix}. \tag{1.98}$$

On a alors $B = P - A = \begin{bmatrix} -1 & 1 \\ 1 & -1 \end{bmatrix}$. Les valeurs propres de B sont 0 et -2 et on a donc $\rho(B) = 2 > 1$. La suite $(\boldsymbol{e}^{(k)})_{k\in\mathbb{N}}$ définie par $\boldsymbol{e}^{(k)} = B^k\boldsymbol{e}^{(0)}$ n'est donc en général pas convergente. En effet, si $\boldsymbol{e}^{(0)} = a\boldsymbol{u}_1 + b\boldsymbol{u}_2$, où $\boldsymbol{u}_1 = \begin{bmatrix} 1 \\ -1 \end{bmatrix}$ est vecteur propre de B associé à la valeur propre $\lambda = -2$, on a $\boldsymbol{e}^{(k)} = (-2)^k a$ et donc $|\boldsymbol{e}^{(k)}| \to +\infty$ lorsque $k \to \infty$ dès que $a \neq 0$. Cette première idée n'est donc pas si bonne…

La méthode de Richardson

Affinons un peu et prenons maintenant $P = \beta \mathrm{Id}$, avec $\beta \in \mathbb{R}$. On a dans ce cas $P - A = \beta \mathrm{Id} - A$ et $B = \mathrm{Id} - \frac{1}{\beta}A = \mathrm{Id} - \alpha A$ avec $\alpha = \frac{1}{\beta}$. Les valeurs propres de B sont de la forme $1 - \alpha\lambda$, où λ est valeur propre de A. Pour la matrice A définie par (1.98), les valeurs propres de A sont 1 et 3, et les valeurs propres de

$$B = \begin{bmatrix} 1-2\alpha & \alpha \\ \alpha & 1-2\alpha, \end{bmatrix}$$

sont $1-\alpha$ et $1-3\alpha$. Le rayon spectral de la matrice B, qui dépend de α est donc $\rho(B) = \max(|1-\alpha|, |1-3\alpha|)$, qu'on représente sur la figure ci-dessous. La méthode itérative s'écrit

$$\begin{aligned} &\boldsymbol{x}^{(0)} \in \mathbb{R}^n \text{ donné}, \\ &\boldsymbol{x}^{(k+1)} = B\boldsymbol{x}^{(k)} + \boldsymbol{c}, \text{ avec } \boldsymbol{c} = \alpha \boldsymbol{b}. \end{aligned} \tag{1.99}$$

Pour que la méthode converge, il faut et il suffit que $\rho(B) < 1$, c.à.d. $3\alpha - 1 < 1$, donc $\alpha < \frac{2}{3}$. On voit que le choix $\alpha = 1$ qu'on avait fait au départ n'était pas bon. Mais on peut aussi calculer le meilleur coefficient α pour avoir la meilleure convergence possible : c'est la valeur de α qui minimise le rayon spectral ρ ; il est atteint pour $1 - \alpha = 3\alpha - 1$, ce qui donne $\alpha = \frac{1}{2}$. Cette méthode est connue sous le nom de *méthode de Richardson*[8]. Elle est souvent écrite sous la forme :

$$\begin{aligned} &\boldsymbol{x}^{(0)} \in \mathbb{R}^n \text{ donné}, \\ &\boldsymbol{x}^{(k+1)} = \boldsymbol{x}^{(k)} + \alpha \boldsymbol{r}^{(k)}, \end{aligned}$$

où $\boldsymbol{r}^{(k)} = \boldsymbol{b} - A\boldsymbol{x}^{(k)}$ est le résidu. On vérifie facilement que cette forme est équivalente à la forme (1.99) qu'on vient d'étudier.

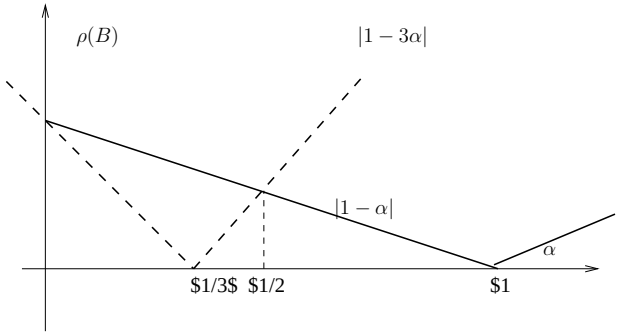

FIGURE 1.4: Rayon spectral de la matrice B de Richardson en fonction du coefficient α.

La méthode de Jacobi

Dans le cas de l'exemple de la matrice A donné par (1.98), la méthode de Richardson avec le coefficient optimal $\alpha = \frac{1}{2}$ revient à prendre comme décomposition de $A = P + A - P$ avec comme matrice $P = D$, où D est la

8. Lewis Fry Richardson, (1881-1953) est un mathématician, physicien, météorologue et psychologue qui a introduit les méthodes mathématiques pour les prévisions météorologiques. Il est également connu pour ses travaux sur les fractals. C'était un pacifiste qui a abandonné ses travaux de météorologie en raison de leur utillisation par l'armée de l'air, pour se tourner vers l'étude des raisons des guerres et de leur prévention.

matrice diagonale dont les coefficients sont les coefficients situés sur la diagonale de A. La *méthode de Jacobi*[9] consiste justement à prendre $P = D$, et ce même si la diagonale de A n'est pas constante.
Elle n'est équivalente à la méthode de Richardson avec coefficient optimal que dans le cas où la diagonale est constante ; c'est le cas de l'exemple (1.98), et donc dans ce cas la méthode de Jacobi s'écrit

$$\begin{aligned} &\boldsymbol{x}^{(0)} = \begin{bmatrix} x_1^{(0)} \\ x_2^{(0)} \end{bmatrix} \in \mathbb{R}^2 \text{ donné ,} \\ &\boldsymbol{x}^{(k+1)} = \begin{bmatrix} x_1^{(k+1)} \\ x_2^{(k+1)} \end{bmatrix} = B_J \boldsymbol{x}^{(k)} + \boldsymbol{c}, \text{ avec } B_J = \begin{bmatrix} 0 & \frac{1}{2} \\ \frac{1}{2} & 0, \end{bmatrix} \text{ et } \boldsymbol{c} = \frac{1}{2}\boldsymbol{b}. \end{aligned} \tag{1.100}$$

Dans le cas d'une matrice A générale, on décompose A sous la forme $A = D - E - F$, où D représente la diagonale de la matrice A, $(-E)$ la partie triangulaire inférieure et $(-F)$ la partie triangulaire supérieure :

$$D = \begin{bmatrix} a_{1,1} & 0 & \dots & 0 \\ 0 & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & & 0 & a_{n,n} \end{bmatrix}, \; -E = \begin{bmatrix} 0 & 0 & \dots & 0 \\ a_{2,1} & \ddots & & \vdots \\ \vdots & \ddots & \ddots & 0 \\ a_{n,1} & \dots & a_{n-1,n} & 0 \end{bmatrix} \text{ et } -F = \begin{bmatrix} 0 & a_{1,2} & \dots & a_{1,n} \\ \vdots & \ddots & & \vdots \\ \vdots & \ddots & \ddots & a_{n,n-1} \\ 0 & \dots & 0 & -0 \end{bmatrix}. \tag{1.101}$$

La méthode de Jacobi s'écrit donc :

$$\begin{cases} \boldsymbol{x}^{(0)} \in \mathbb{R}^n \\ D\boldsymbol{x}^{(k+1)} = (E+F)\boldsymbol{x}^{(k)} + \boldsymbol{b}. \end{cases} \tag{1.102}$$

Lorsqu'on écrit la méthode de Jacobi comme sous la forme (1.92) on a $B = D^{-1}(E+F)$; on notera B_J cette matrice :

$$B_J = \begin{bmatrix} 0 & -\frac{a_{1,2}}{a_{1,1}} & \dots & -\frac{-a_{1,n}}{a_{1,1}} \\ -\frac{a_{2,1}}{a_{2,2}} & \ddots & & -\frac{-a_{2,n}}{a_{2,2}} \\ \vdots & \ddots & \ddots & \vdots \\ -\frac{a_{n,1}}{a_{n,n}} & \dots & -\frac{-a_{n-1,n}}{a_{n,n}} & 0 \end{bmatrix}.$$

La méthode de Jacobi s'écrit aussi :

$$\begin{cases} \boldsymbol{x}^{(0)} \in \mathbb{R}^n \\ a_{i,i} x_i^{(k+1)} = -\sum_{j<i} a_{i,j} x_j^{(k)} - \sum_{j>i} a_{i,j} x_j^{(k)} + b_i \;\; i = 1, \dots, n. \end{cases} \tag{1.103}$$

La méthode de Gauss-Seidel

Dans l'écriture (1.103) de la méthode de Jacobi, on pourrait remplacer les composantes $x_j^{(k)}$ dans la somme pour $j < i$ par les composantes $x_j^{(k+1)}$, puisqu'elles sont déjà calculées au moment où l'on calcule $x_i^{(k+1)}$. C'est l"idée de la méthode de Gauss-Seidel[10] qui consiste à utiliser le calcul des composantes de l'itéré $(k+1)$ dès qu'il est effectué. Par exemple, pour calculer la deuxième composante $x_2^{(k+1)}$ du vecteur $x^{(k+1)}$, on pourrait employer la

9. Carl G. J. Jacobi, (1804 - 1851), mathématicien allemand. Issu d'une famille juive, il étudie à l'Université de Berlin, où il obtient son doctorat à 21 ans. Sa thèse est une discussion analytique de la théorie des fractions. En 1829, il devient professeur de mathématique à l'Université de Königsberg, et ce jusqu'en 1842. Il fait une dépression, et voyage en Italie en 1843. À son retour, il déménage à Berlin où il sera pensionnaire royal jusqu'à sa mort. Sa lettre du 2 juillet 1830 adressée à Legendre est restée célèbre pour la phrase suivante, qui a fait couler beaucoup d'encre : "M. Fourier avait l'opinion que le but principal des mathématiques était l'utilité publique et l'explication des phénomènes naturels ; mais un philosophe comme lui aurait dû savoir que le but unique de la science, c'est l'honneur de l'esprit humain, et que sous ce titre, une question de nombres vaut autant qu'une question du système du monde." C'est une question toujours en discussion. . . .

10. Philipp Ludwig von Seidel (Zweibrücken, Allemagne 1821 – Munich, 13 August 1896) mathématicien allemand dont il est dit qu'il a découvert en 1847 le concept crucial de la convergence uniforme en étudiant une démonstration incorrecte de Cauchy.

"nouvelle" valeur $x_1^{(k+1)}$ qu'on vient de calculer plutôt que la valeur $x_1^{(k)}$ comme dans (1.103) ; de même, dans le calcul de $x_3^{(k+1)}$, on pourrait employer les "nouvelles" valeurs $x_1^{(k+1)}$ et $x_2^{(k+1)}$ plutôt que les valeurs $x_1^{(k)}$ et $x_2^{(k)}$. Cette idée nous suggère de remplacer dans (1.103) $x_j^{(k)}$ par $x_j^{(k+1)}$ si $j < i$. On obtient donc l'algorithme suivant :

$$\left\{ \begin{array}{l} \boldsymbol{x}^{(0)} \in \mathbb{R}^n \\ a_{i,i} x_i^{(k+1)} = -\sum_{j<i} a_{i,j} x_j^{(k+1)} - \sum_{i<j} a_{i,j} x_j^{(k)} + b_i, \quad i = 1, \ldots, n. \end{array} \right. \tag{1.104}$$

La méthode de Gauss–Seidel s'écrit donc sous la forme $P\boldsymbol{x}^{(k+1)} = (P - A)\boldsymbol{x}^{(k)} + \boldsymbol{b}$, avec $P = D - E$ et $P - A = F$:

$$\left\{ \begin{array}{l} \boldsymbol{x}_0 \in \mathbb{R}^n \\ (D - E)\boldsymbol{x}^{(k+1)} = F\boldsymbol{x}^{(k)} + b. \end{array} \right. \tag{1.105}$$

Si l'on écrit la méthode de Gauss–Seidel sous la forme $\boldsymbol{x}^{(k+1)} = B\boldsymbol{x}^{(k)} + \boldsymbol{c}$, on voit assez vite que $B = (D - E)^{-1}F$; on notera B_{GS} cette matrice, dite matrice de Gauss-Seidel.
Ecrivons la méthode de Gauss-Seidel dans le cas de la matrice A donnée par (1.98) : on a dans ce cas $P = D - E = \begin{bmatrix} 2 & 0 \\ -1 & 2 \end{bmatrix}$, $F = \begin{bmatrix} 0 & 1 \\ 0 & 0, \end{bmatrix}$. L'algorithme de Gauss-Seidel s'écrit donc :

$$\begin{array}{l} \boldsymbol{x}^{(0)} = \begin{bmatrix} x_1^{(0)} \\ x_2^{(0)} \end{bmatrix} \in \mathbb{R}^2 \text{ donné} , \\ \boldsymbol{x}^{(k+1)} = \begin{bmatrix} x_1^{(k+1)} \\ x_2^{(k+1)} \end{bmatrix} = B_{GS}\boldsymbol{x}^{(k)} + \boldsymbol{c}, \text{ avec } B_{GS} = \begin{bmatrix} 0 & 0 \\ 0 & \frac{1}{4} \end{bmatrix} \text{ et } \boldsymbol{c} = \begin{bmatrix} \frac{1}{2} & 0 \\ \frac{1}{4} & \frac{1}{2} \end{bmatrix} \boldsymbol{b}. \end{array} \tag{1.106}$$

On a donc $\rho(B_{GS}) = \frac{1}{4}$. Sur cet exemple la méthode de Gauss-Seidel converge donc beaucoup plus vite que la méthode de Jacobi : Asymptotiquement, l'erreur est divisée par 4 à chaque itération au lieu de 2 pour la méthode de Jacobi. On peut montrer que c'est le cas pour toutes les matrices tridiagonales, comme c'est énoncé dans le théorème suivant :

Théorème 1.51 (Comparaison de Jacobi et Gauss-Seidel pour les matrices tridiagonales)**.** *On considère une matrice $A \in \mathcal{M}_n(\mathbb{R})$ tridiagonale, c.à.d. telle que $a_{i,j} = 0$ si $|i - j| > 1$; soient B_{GS} et B_{J} les matrices d'itération respectives des méthodes de Gauss-Seidel et Jacobi, alors :*

$$\rho(B_{\mathrm{GS}}) = (\rho(B_{\mathrm{J}}))^2.$$

Pour les matrices tridiagonales, la méthode de Gauss–Seidel converge (ou diverge) donc plus vite que celle de Jacobi.

La démonstration de ce résultat se fait en montrant que dans le cas tridiagonal, λ est valeur propre de la matrice d'itération de Jacobi si et seulement si λ^2 est valeur propre de la matrice d'itération de Gauss-Seidel. Elle est laissée à titre d'exercice.

Méthodes SOR et SSOR

L'idée de la méthode de sur-relaxation (SOR = Successive Over Relaxation) est d'utiliser la méthode de Gauss-Seidel pour calculer un itéré intermédiaire $\tilde{x}^{(k+1)}$ qu'on "relaxe" ensuite pour améliorer la vitesse de convergence de la méthode. On se donne $0 < \omega < 2$, et on modifie l'algorithme de Gauss–Seidel de la manière suivante :

$$\left\{ \begin{array}{l} x_0 \in \mathbb{R}^n \\ a_{i,i}\tilde{x}_i^{(k+1)} = -\displaystyle\sum_{j<i} a_{i,j} x_j^{(k+1)} - \sum_{i<j} a_{i,j} x_j^{(k)} + b_i \\ x_i^{(k+1)} = \omega\tilde{x}_i^{(k+1)} + (1 - \omega)x_i^{(k)}, \quad i = 1, \ldots, n. \end{array} \right. \tag{1.107}$$

(Pour $\omega = 1$ on retrouve la méthode de Gauss–Seidel.)
L'algorithme ci-dessus peut aussi s'écrire (en multipliant par $a_{i,i}$ la ligne 3 de l'algorithme (1.107)) :

$$\left\{ \begin{array}{l} x^{(0)} \in \mathbb{R}^n \\ a_{i,i}x_i^{(k+1)} = \omega \left[-\sum_{j<i} a_{i,j}x_j^{(k+1)} - \sum_{j>i} a_{i,j}x_j^{(k)} + b_i \right] \\ \qquad\qquad +(1-\omega)a_{i,i}x_i^{(k)}. \end{array} \right. \tag{1.108}$$

On obtient donc

$$(D - \omega E)x^{(k+1)} = \omega F x^{(k)} + \omega b + (1-\omega)Dx^{(k)}.$$

La matrice d'itération de l'algorithme SOR est donc

$$B_\omega = \left(\frac{D}{\omega} - E\right)^{-1} (F + \left(\frac{1-\omega}{\omega}\right) D) = P^{-1}N, \text{ avec } P = \frac{D}{\omega} - E \text{ et } N = F + \left(\frac{1-\omega}{\omega}\right) D.$$

Il est facile de vérifier que $A = P - N$.

Proposition 1.52 (Condition nécessaire de convergence de la méthode SOR)**.**
Soit $A \in \mathcal{M}_n(\mathbb{R})$ et soiet D, E et F les matrices définies par (1.101)*; on a donc $A = D - E - F$. Soit B_ω la matrice d'itération de la méthode SOR (et de la méthode de Gauss–Seidel pour $\omega = 1$) définie par :*

$$B_\omega = \left(\frac{D}{\omega} - E\right)^{-1} \left(F + \frac{1-\omega}{\omega} D\right), \quad \omega \neq 0.$$

Si $\rho(B_\omega) < 1$ alors $0 < \omega < 2$.

DÉMONSTRATION – Calculons $\det(B_\omega)$. Par définition,

$$B_\omega = P^{-1}N, \text{ avec } P = \frac{1}{\omega}D - E \text{ et } N = F + \frac{1-\omega}{\omega}D.$$

Donc $\det(B_\omega) = (\det(P))^{-1}\det(N)$. Comme P et N sont des matrices triangulaires, leurs déterminants sont les produits ceofficients diagonaux (voir la remarque 1.59 page 106). On a donc :

$$\det(B_\omega) = \frac{(\frac{1-\omega}{\omega})^n \det(D)}{(\frac{1}{\omega})^n \det(D)} = (1-\omega)^n.$$

Or le déterminant d'une matrice est aussi le produit des valeurs propres de cette matrice (comptées avec leur multiplicités algébriques), dont les valeurs absolues sont toutes inférieures au rayon spectral. On a donc : $|\det(B_\omega)| = |(1-\omega)^n| \leq (\rho(B_\omega))^n$, d'où le résultat. ∎

On a un résultat de convergence de la méthode SOR (et donc également de Gauss-Seidel) dans le cas où A est symétrique définie positive, grâce au lemme suivant :

Lemme 1.53 (Condition suffisante de convergence pour la suite définie par (1.92))**.** *Soit $A \in \mathcal{M}_n(\mathbb{R})$ une matrice symétrique définie positive, et soient P et $N \in \mathcal{M}_n(\mathbb{R})$ telles que $A = P - N$ et P est inversible. Si la matrice $P^t + N$ est symétrique définie positive alors $\rho(P^{-1}N) = \rho(B) < 1$, et donc la suite définie par* (1.92) *converge.*

DÉMONSTRATION – On rappelle (voir le corollaire (1.36) page 69) que si $B \in \mathcal{M}_n(\mathbb{R})$, et si $\|\cdot\|$ est une norme induite sur $\mathcal{M}_n(\mathbb{R})$ par une norme sur $\mathbb{R}^n$, on a toujours $\rho(B) \leq \|B\|$. On va donc chercher une norme sur $\mathbb{R}^n$, notée $\|\cdot\|_*$ telle que

$$\|P^{-1}N\|_* = \max\{\|P^{-1}N\boldsymbol{x}\|_*,\ \boldsymbol{x} \in \mathbb{R}^n,\ \|\boldsymbol{x}\|_* = 1\} < 1,$$

(où on désigne encore par $\|.\|_*$ la norme induite sur $\mathcal{M}_n(\mathbb{R})$) ou encore :

$$\|P^{-1}N\boldsymbol{x}\|_* < \|x\|_*,\ \forall \boldsymbol{x} \in \mathbb{R}^n,\ \boldsymbol{x} \neq 0. \tag{1.109}$$

On définit la norme $\|\cdot\|_*$ par $\|\boldsymbol{x}\|_* = \sqrt{A\boldsymbol{x}\cdot\boldsymbol{x}}$, pour tout $\boldsymbol{x} \in \mathbb{R}^n$. Comme A est symétrique définie positive, $\|\cdot\|_*$ est bien une norme sur $\mathbb{R}^n$, induite par le produit scalaire $(\boldsymbol{x}|\boldsymbol{y})_A = A\boldsymbol{x}\cdot\boldsymbol{y}$. On va montrer que la propriété (1.109) est vérifiée par cette norme. Soit $\boldsymbol{x} \in \mathbb{R}^n$, $\boldsymbol{x} \neq 0$, on a : $\|P^{-1}N\boldsymbol{x}\|_*^2 = AP^{-1}N\boldsymbol{x}\cdot P^{-1}N\boldsymbol{x}$. Or $N = P - A$, et donc : $\|P^{-1}N\boldsymbol{x}\|_*^2 = A(\mathrm{Id} - P^{-1}A)\boldsymbol{x}\cdot(Id - P^{-1}A)\boldsymbol{x}$. Soit $\boldsymbol{y} = P^{-1}A\boldsymbol{x}$; remarquons que $\boldsymbol{y} \neq 0$ car $\boldsymbol{x} \neq 0$ et $P^{-1}A$ est inversible. Exprimons $\|P^{-1}N\boldsymbol{x}\|_*^2$ à l'aide de $\boldsymbol{y}$.

$$\|P^{-1}N\boldsymbol{x}\|_*^2 = A(\boldsymbol{x}-\boldsymbol{y})\cdot(\boldsymbol{x}-\boldsymbol{y}) = A\boldsymbol{x}\cdot\boldsymbol{x} - 2A\boldsymbol{x}\cdot\boldsymbol{y} + A\boldsymbol{y}\cdot\boldsymbol{y} = \|\boldsymbol{x}\|_*^2 - 2A\boldsymbol{x}\cdot\boldsymbol{y} + A\boldsymbol{y}\cdot\boldsymbol{y}.$$

Pour que $\|P^{-1}N\boldsymbol{x}\|_*^2 < \|\boldsymbol{x}\|_*^2$ (et par suite $\rho(P^{-1}N) < 1$), il suffit donc de montrer que $-2A\boldsymbol{x}\cdot\boldsymbol{y} + A\boldsymbol{y}\cdot\boldsymbol{y} < 0$. Or, comme $P\boldsymbol{y} = A\boldsymbol{x}$, on a : $-2A\boldsymbol{x}\cdot\boldsymbol{y} + A\boldsymbol{y}\cdot\boldsymbol{y} = -2P\boldsymbol{y}\cdot\boldsymbol{y} + A\boldsymbol{y}\cdot\boldsymbol{y}$. En écrivant : $P\boldsymbol{y}\cdot\boldsymbol{y} = \boldsymbol{y}\cdot P^t\boldsymbol{y} = P^t\boldsymbol{y}\cdot\boldsymbol{y}$, on obtient donc que : $-2A\boldsymbol{x}\cdot\boldsymbol{y} + A\boldsymbol{y}\cdot\boldsymbol{y} = (-P - P^t + A)\boldsymbol{y}\cdot\boldsymbol{y}$, et comme $A = P - N$ on obtient $-2A\boldsymbol{x}\cdot\boldsymbol{y} + A\boldsymbol{y}\cdot\boldsymbol{y} = -(P^t + N)\boldsymbol{y}\cdot\boldsymbol{y}$. Comme $P^t + N$ est symétrique définie positive par hypothèse et que $\boldsymbol{y} \neq \mathbf{0}$, on en déduit que $-2A\boldsymbol{x}\cdot\boldsymbol{y} + A\boldsymbol{y}\cdot\boldsymbol{y} < 0$, ce qui termine la démonstration. ∎

Théorème 1.54 (CNS de convergence de la méthode SOR pour les matrices s.d.p.)**.**
Soit $A \in \mathcal{M}_n(\mathbb{R})$ une matrice symétrique définie positive, et soient D, E et F les matrices définies par (1.101)*; on a donc $A = D - E - F$. Soit B_ω la matrice d'itération de la méthode SOR (et de la méthode de Gauss–Seidel pour $\omega = 1$) définie par :*

$$B_\omega = \left(\frac{D}{\omega} - E\right)^{-1}\left(F + \frac{1-\omega}{\omega}D\right),\quad \omega \neq 0.$$

Alors :

$$\rho(B_\omega) < 1 \text{ si et seulement si } 0 < \omega < 2.$$

En particulier, si A est une matrice symétrique définie positive, la méthode de Gauss–Seidel converge.

DÉMONSTRATION – On sait par la proposition 1.52 que si $\rho(B_\omega) < 1$ alors $0 < \omega < 2$. Supposons maintenant que A est une matrice symétrique définie positive, que $0 < \omega < 2$ et montrons que $\rho(B_\omega) < 1$. Par le lemme 1.53 page 103, il suffit pour cela de montrer que $P^t + N$ est une matrice symétrique définie positive. Or,

$$P^t = \left(\frac{D}{\omega} - E\right)^t = \frac{D}{\omega} - F,$$

$$P^t + N = \frac{D}{\omega} - F + F + \frac{1-\omega}{\omega}D = \frac{2-\omega}{\omega}D.$$

La matrice $P^t + N$ est donc bien symétrique définie positive. ∎

Remarque 1.55 (Comparaison Gauss–Seidel/Jacobi)**.** *On a vu (théorème 1.54) que si A est une matrice symétrique définie positive, la méthode de Gauss–Seidel converge. Par contre, même dans le cas où A est symétrique définie positive, il existe des cas où la méthode de Jacobi ne converge pas, voir à ce sujet l'exercice 54 page 109.*

Remarquons que le résultat de convergence des méthodes itératives donné par le théorème précédent n'est que partiel, puisqu'il ne concerne que les matrices symétriques définies positives et que les méthodes Gauss-Seidel et SOR. On a aussi un résultat de convergence de la méthode de Jacobi pour les matrices à diagonale dominante stricte, voir exercice 59 page 111, et un résultat de comparaison des méthodes pour les matrices tridiagonales par blocs, voir le théorème 1.56 donné ci-après. Dans la pratique, il faudra souvent compter sur sa bonne étoile…

Estimation du coefficient de relaxation optimal de SOR La question est ici d'estimer le coefficient de relaxation ω optimal dans la méthode SOR, c.à.d. le coefficient $\omega_0 \in]0,2[$ (condition nécessaire pour que la méthode SOR converge, voir théorème 1.54) tel que

$$\rho(B_{\omega_0}) \leq \rho(B_\omega), \forall \omega \in]0,2[.$$

Ce coefficient ω_0 donnera la meilleure convergence possible pour SOR. On sait le faire dans le cas assez restrictif des matrices tridiagonales (ou tridiagonales par blocs, voir paragraphe suivant). On ne fait ici qu'énoncer le résultat dont la démonstration est donnée dans le livre de Ph.Ciarlet conseillé en début de cours.

Théorème 1.56 (Coefficient optimal, matrice tridiagonale)**.** *On considère une matrice $A \in \mathcal{M}_n(\mathbb{R})$ qui admet une décomposition par blocs définie dans la définition 1.110 page 106 ; on suppose que la matrice A est tridiagonale par blocs, c.à.d. $A_{i,j} = 0$ si $|i-j| > 1$; soient $B_{\rm GS}$ et $B_{\rm J}$ les matrices d'itération respectives des méthodes de Gauss-Seidel et Jacobi. On suppose de plus que toutes les valeurs propres de la matrice d'itération J de la méthode de Jacobi sont réelles et que $\rho(B_{\rm J}) < 1$. Alors le paramètre de relaxation optimal, c.à.d. le paramètre ω_0 tel que $\rho(B_{\omega_0}) = \min\{\rho(B_\omega), \omega \in]0,2[\}$, s'exprime en fonction du rayon spectral $\rho(B_{\rm J})$ de la matrice J par la formule :*

$$\omega_0 = \frac{2}{1+\sqrt{1-\rho(B_{\rm J})^2}} > 1,$$

et on a : $\rho(B_{\omega_0}) = \omega_0 - 1$.

La démonstration de ce résultat repose sur la comparaison des valeurs propres des matrices d'itération. On montre que λ est valeur propre de B_ω si et seulement si

$$(\lambda + \omega - 1)^2 = \lambda \omega \mu^2,$$

où μ est valeur propre de $B_{\rm J}$ (voir [Ciarlet] pour plus de détails).

Remarque 1.57 (Méthode de Jacobi relaxée)**.** *On peut aussi appliquer une procédure de relaxation avec comme méthode iérative "de base" la méthode de Jacobi, voir à ce sujet l'exercice 56 page 110). Cette méthode est toutefois beaucoup moins employée en pratique (car moins efficace) que la méthode SOR.*

Méthode SSOR En "symétrisant" le procédé de la méthode SOR, c.à.d. en effectuant les calculs SOR sur les blocs dans l'ordre 1 à n puis dans l'ordre n à 1, on obtient la méthode de sur-relaxation symétrisée (SSOR = Symmetric Successive Over Relaxation) qui s'écrit dans le formalisme de la méthode I avec

$$B_{SSOR} = \underbrace{\left(\frac{D}{\omega} - F\right)^{-1}\left(E + \frac{1-\omega}{\omega}D\right)}_{\text{calcul dans l'ordre } n\ldots 1}\underbrace{\left(\frac{D}{\omega} - E\right)^{-1}\left(F + \frac{1-\omega}{\omega}D\right)}_{\text{calcul dans l'ordre } 1\ldots n}.$$

1.5.3 Les méthodes par blocs

Décomposition par blocs d'une matrice

Dans de nombreux cas pratiques, les matrices des systèmes linéaires à résoudre ont une structure "par blocs", et on se sert alors de cette structure lors de la résolution par une méthode itérative.

Définition 1.58. *Soit $A \in \mathcal{M}_n(\mathbb{R})$ une matrice inversible ; une décomposition par blocs de A est définie par un entier $S \leq n$, des entiers $(n_i)_{i=1,\dots,S}$ tels que $\sum_{i=1}^{S} n_i = n$, et S^2 matrices $A_{i,j} \in \mathcal{M}_{n_i,n_j}(\mathbb{R})$ (ensemble des matrices rectangulaires à n_i lignes et n_j colonnes, telles que les matrices $A_{i,i}$ soient inversibles pour $i = 1, \dots, S$ et*

$$A = \begin{bmatrix} A_{1,1} & A_{1,2} & \dots & & \dots & A_{1,S} \\ A_{2,1} & \ddots & \ddots & & & \vdots \\ \vdots & \ddots & \ddots & \ddots & & \\ & & \ddots & \ddots & \ddots & \vdots \\ \vdots & & & \ddots & \ddots & A_{S-1,S} \\ A_{S,1} & \dots & & \dots & A_{S,S-1} & A_{S,S} \end{bmatrix} \tag{1.110}$$

Remarque 1.59.

1. *Si $S = n$ et $n_i = 1$ $\forall i \in \{1, \dots, S\}$, chaque bloc est constitué d'un seul coefficient, et on retrouve la structure habituelle d'une matrice. Les méthodes que nous allons décrire maintenant sont alors celles que nous avons vu dans le cas de matrices sans structure particulière.*
2. *Si A est symétrique définie positive, la condition $A_{i,i}$ inversible dans la définition 1.58 est inutile car $A_{i,i}$ est nécessairement symétrique définie positive donc inversible. Pour s'en convaincre, prenons par exemple $i = 1$; soit $y \in \mathbb{R}^{n_1}$, $y \neq 0$ et $x = (y, 0 \dots, 0)^t \in \mathbb{R}^n$. Alors $A_{1,1}y \cdot y = Ax \cdot x > 0$ donc $A_{1,1}$ est symétrique définie positive.*
3. *Si A est une matrice triangulaire par blocs, c.à.d. de la forme* (1.110) *avec $A_{i,j} = 0$ si $j > i$, alors*

$$\det(A) = \prod_{i=1}^{S} \det(A_{i,i}).$$

Par contre si A est décomposée en 2×2 blocs carrés (i.e. tels que $n_i = m_j$, $\forall (i,j) \in \{1,2\}$), on a en général :

$$\det(A) \neq \det(A_{1,1})\det(A_{2,2}) - \det(A_{1,2})\det(A_{2,1}).$$

Méthode de Jacobi

On cherche une matrice P tel que le système $Px = (P - A)x + b$ soit facile à résoudre (on rappelle que c'est un objectif dans la construction d'une méthode itérative). On avait pris pour P une matrice diagonale dans la méthode de Jacobi. La méthode de Jacobi par blocs consiste à prendre pour P la matrice diagonale D formée par les blocs diagonaux de A :

$$D = \begin{bmatrix} A_{1,1} & 0 & \dots & & \dots & 0 \\ 0 & \ddots & \ddots & & & \vdots \\ \vdots & \ddots & \ddots & \ddots & & \\ & & \ddots & \ddots & \ddots & \vdots \\ \vdots & & & \ddots & \ddots & 0 \\ 0 & \dots & & \dots & 0 & A_{S,S} \end{bmatrix}.$$

Dans la matrice ci-dessus, 0 désigne un bloc nul.

On a alors $N = P - A = E + F$, où E et F sont constitués des blocs triangulaires inférieurs et supérieurs de la matrice A :

$$E = \begin{bmatrix} 0 & 0 & \dots & & \dots & 0 \\ -A_{2,1} & \ddots & \ddots & & & \vdots \\ \vdots & \ddots & \ddots & \ddots & & \\ & & \ddots & \ddots & \ddots & \vdots \\ \vdots & & & \ddots & \ddots & 0 \\ -A_{S,1} & \dots & & \dots & -A_{S,S-1} & 0 \end{bmatrix}, \; F = \begin{bmatrix} 0 & -A_{1,2} & \dots & & \dots & -A_{1,S} \\ 0 & \ddots & \ddots & & & \vdots \\ \vdots & \ddots & \ddots & \ddots & & \\ & & \ddots & \ddots & \ddots & \vdots \\ \vdots & & & \ddots & \ddots & -A_{S-1,S} \\ 0 & \dots & & \dots & 0 & 0 \end{bmatrix}.$$

On a bien $A = P - N$ et avec D, E et F définies comme ci-dessus, la méthode de Jacobi s'écrit :

$$\begin{cases} x^{(0)} \in \mathbb{R}^n \\ Dx^{(k+1)} = (E+F)x^{(k)} + b. \end{cases} \tag{1.111}$$

Lorsqu'on écrit la méthode de Jacobi comme sous la forme (1.92) on a $B = D^{-1}(E+F)$; on notera J cette matrice. En introduisant la décomposition par blocs de x, solution recherchée de (1.1), c.à.d. : $x = [x_1, \dots, x_S]^t$, où $x_i \in \mathbb{R}^{n_i}$, on peut aussi écrire la méthode de Jacobi sous la forme :

$$\begin{cases} x_0 \in \mathbb{R}^n \\ A_{i,i}x_i^{(k+1)} = -\displaystyle\sum_{j<i} A_{i,j}x_j^{(k)} - \sum_{j>i} A_{i,j}x_j^{(k)} + b_i \;\; i = 1, \dots, S. \end{cases} \tag{1.112}$$

Si $S = n$ et $n_i = 1 \; \forall i \in \{1, \dots, S\}$, chaque bloc est constitué d'un seul coefficient, et on obtient la méthode de Jacobi par points (aussi appelée méthode de Jacobi), qui s'écrit donc :

$$\begin{cases} x_0 \in \mathbb{R}^n \\ a_{i,i}x_i^{(k+1)} = -\displaystyle\sum_{j<i} a_{i,j}x_j^{(k)} - \sum_{j>i} a_{i,j}x_j^{(k)} + b_i \;\; i = 1, \dots, n. \end{cases} \tag{1.113}$$

Méthode de Gauss-Seidel

La même procédure que dans le cas $S = n$ et $n_i = 1$ donne :

$$\begin{cases} x^{(0)} \in \mathbb{R}^n \\ A_{i,i}x_i^{(k+1)} = -\sum_{j<i} A_{i,j}x_j^{(k+1)} - \sum_{i<j} A_{i,j}x_j^{(k)} + b_i, \quad i = 1, \dots, S. \end{cases} \tag{1.114}$$

La méthode de Gauss–Seidel s'écrit donc sous forme la forme $Px^{(k+1)} = (P-A)x^{(k)} + b$, $P = D - E$ et $P - A = F$:

$$\begin{cases} x_0 \in \mathbb{R}^n \\ (D-E)x^{(k+1)} = Fx^{(k)} + b. \end{cases} \tag{1.115}$$

Si l'on écrit la méthode de Gauss–Seidel sous la forme $x^{(k+1)} = Bx^{(k)} + c$, on voit assez vite que $B = (D - E)^{-1}F$; on notera B_{GS} cette matrice, dite matrice de Gauss-Seidel.

Méthodes SOR et SSOR

La méthode SOR s'écrit aussi par blocs : on se donne $0 < \omega < 2$, et on modifie l'algorithme de Gauss–Seidel de la manière suivante :

$$\begin{cases} x_0 \in \mathbb{R}^n \\ A_{i,i}\tilde{x}_i^{(k+1)} = -\displaystyle\sum_{j<i} A_{i,j}x_j^{(k+1)} - \sum_{i<j} A_{i,j}x_j^{(k)} + b_i \\ x_i^{(k+1)} = \omega\tilde{x}_i^{(k+1)} + (1-\omega)x_i^{(k)}, \quad i = 1, \dots, S. \end{cases} \tag{1.116}$$

(Pour $\omega = 1$ on retrouve la méthode de Gauss–Seidel.)
L'algorithme ci-dessus peut aussi s'écrire (en multipliant par $A_{i,i}$ la ligne 3 de l'algorithme (1.107)) :

$$\left\{ \begin{array}{l} x^{(0)} \in \mathbb{R}^n \\ A_{i,i} x_i^{(k+1)} = \omega \left[-\sum_{j<i} A_{i,j} x_j^{(k+1)} - \sum_{j>i} A_{i,j} x_j^{(k)} + b_i \right] \\ \qquad\qquad +(1-\omega) A_{i,i} x_i^{(k)}. \end{array} \right. \tag{1.117}$$

On obtient donc

$$(D - \omega E) x^{(k+1)} = \omega F x^{(k)} + \omega b + (1-\omega) D x^{(k)}.$$

L'algorithme SOR s'écrit donc comme une méthode II avec

$$P = \frac{D}{\omega} - E \text{ et } N = F + \left(\frac{1-\omega}{\omega} \right) D.$$

Il est facile de vérifier que $A = P - N$.
L'algorithme SOR s'écrit aussi comme une méthode I avec

$$B = \left(\frac{D}{\omega} - E \right)^{-1} (F + \left(\frac{1-\omega}{\omega} \right) D).$$

Remarque 1.60 (Méthode de Jacobi relaxée)**.** *On peut aussi appliquer une procédure de relaxation avec comme méthode iérative "de base" la méthode de Jacobi, voir à ce sujet l'exercice 56 page 110). Cette méthode est toutefois beaucoup moins employée en pratique (car moins efficace) que la méthode SOR.*

En "symétrisant" le procédé de la méthode SOR, c.à.d. en effectuant les calculs SOR sur les blocs dans l'ordre 1 à n puis dans l'ordre n à 1, on obtient la méthode de sur-relaxation symétrisée (SSOR = Symmetric Successive Over Relaxation) qui s'écrit dans le formalisme de la méthode I avec

$$B = \underbrace{\left(\frac{D}{\omega} - F \right)^{-1} \left(E + \frac{1-\omega}{\omega} D \right)}_{\text{calcul dans l'ordre } S \ldots 1} \underbrace{\left(\frac{D}{\omega} - E \right)^{-1} \left(F + \frac{1-\omega}{\omega} D \right)}_{\text{calcul dans l'ordre } 1 \ldots S}.$$