Chapitre premier
Notions fondamentales

Dans ce Chapitre, nous allons faire les rappels nécessaires pour la suite. Cependant, les bases de l'algèbre linéaire seront supposées acquises. Pour plus de détails, se reporter, par exemple, à [4].

Les premiers chapitres sont consacrés à l'étude de méthodes itératives pour la résolution des systèmes d'équations linéaires. Il nous faudra donc en étudier la convergence et mesurer l'écart entre un itéré et la solution exacte. Nous aurons donc besoin de la notion de norme (de vecteur et de matrice). Ces notions seront également utiles dans les autres chapitres. En pratique, il n'est naturellement pas question d'effectuer une infinité d'itérations. Nous aurons donc besoin de tests d'arrêt.
Une autre notion importante est celle de préconditionnement qui, non seulement, permet de réduire les effets de l'arithmétique à précision finie des ordinateurs, mais joue également un rôle important dans la vitesse de convergence de certaines méthodes itératives.
Nous traiterons ensuite des méthodes itératives pour calculer les valeurs propres et les vecteurs propres d'une matrice. Les définitions et résultats généraux sur ce sujet sont supposés connus.
On passera ensuite aux méthodes de résolution des équations non linéaires. Ce chapitre ne fait appel à aucune connaissance particulière. Il en est de même du suivant qui introduit la notion de fractal.
Les méthodes itératives convergent souvent lentement. On transforme alors la suite fournie en une nouvelle suite qui, sous certaines conditions, converge plus vite vers la même limite. Le dernier chapitre traite de ces méthodes et ne requiert aucune connaissance préalable. Il en est de même des deux derniers chapitres.

1. Normes de vecteurs et de matrices

Soit E un espace vectoriel sur $\mathbb{C}$. On appelle *norme* toute application de E dans $\mathbb{R}$ qui, à tout $x \in E$, associe le nombre réel $\|x\|$ (appelé norme de x) qui vérifie les trois conditions suivantes

1. $\|x\| \geq 0$ et $\|x\| = 0$ si et seulement si $x = 0 \in E$,
2. $\|\lambda x\| = |\lambda| \cdot \|x\|, \quad \forall \lambda \in \mathbb{C}$,
3. $\|x+y\| \leq \|x\| + \|y\|, \quad \forall x, y \in E$.

Comme c'est le cas pour la valeur absolue, une norme sert à mesurer la distance $\|x-y\|$ entre deux éléments. Sur un espace vectoriel, il se peut que l'on puisse définir plusieurs normes. Cependant, si l'espace est de dimension finie, toutes les normes sont équivalentes c'est-à-dire qu'il existera des inégalités entre elles. Par conséquent si deux éléments sont

voisins au sens d'une certaine norme (et en particulier si une suite converge pour une certaine norme), il en sera de même pour une autre norme.

Prenons maintenant le cas $E = \mathbb{C}^n$. Soit $x = (x_1, \ldots, x_n)^T$ un vecteur. Les quantités suivantes s'appellent *normes de Hölder* d'indice k

$$\begin{aligned} \|x\|_k &= \left(|x_1|^k + \cdots + |x_n|^k\right)^{1/k}, \qquad k = 1, 2, \ldots, \\ \|x\|_\infty &= \max_{1 \le i \le n} |x_i|. \end{aligned}$$

Parmi ces normes, les plus utilisées sont $\|x\|_1$, $\|x\|_\infty$ et $\|x\|_2$. On les désigne souvent respectivement sous les termes de norme l_1, l_∞ et l_2. Cette dernière représente la longueur du vecteur x au sens euclidien du terme ; elle s'appelle *norme euclidienne*.

Une norme de matrice peut être définie à partir d'une norme pour les vecteurs (mais cela n'est pas obligatoire). Soit $A \in \mathbb{C}^{n \times m}$ une matrice. La quantité

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

est une norme pour la matrice A. Puisque l'on peut, dans cette définition, remplacer x par αx, où α est un scalaire, on peut toujours choisir x de norme 1 et l'on a donc également

$$\|A\| = \sup_{\|x\|=1} \|Ax\|.$$

Ces normes de matrice vérifient les trois propriétés des normes mais, en plus, il existe deux autres propriétés qui nous serons très utiles par la suite

$$\|Ax\| \le \|A\| \cdot \|x\|$$

et

$$\|AB\| \le \|A\| \cdot \|B\|.$$

On appelle *multiplicative* toute norme vérifiant cette dernière inégalité.
Les normes de matrices les plus utilisées sont celles qui sont reliées à une norme de Hölder pour les vecteurs, c'est-à-dire

$$\|A\|_k = \sup_{x \neq 0} \frac{\|Ax\|_k}{\|x\|_k}$$

où $\|\cdot\|_k$ est la norme de vecteur de Hölder d'indice k.
Les normes de matrices semblent être difficiles à calculer en pratique puisqu'elles font intervenir une borne supérieure. Cependant, on connaît leurs expressions exactes dans trois cas pour les normes de Hölder

$$\begin{aligned} \|A\|_1 &= \max_{1 \le j \le m} \sum_{i=1}^{n} |a_{ij}|, \\ \|A\|_\infty &= \max_{1 \le i \le n} \sum_{j=1}^{m} |a_{ij}|, \\ \|A\|_2 &= \sqrt{\rho(A^T A)}, \end{aligned}$$

où $\rho(A^TA)$ désigne le rayon spectral de la matrice $A^TA \in \mathbb{C}^{m\times m}$, c'est-à-dire sa plus grande valeur propre puisque, A^TA étant symétrique définie positive, celles-ci sont réelles et positives. Cette dernière norme s'appelle aussi *norme spectrale*.

À partir de maintenant, les matrices considérées sont carrées ($m = n$).
On démontre que, quelle que soit la norme,

$$\rho(A) \leq \|A\|.$$

Si A est symétrique, $\|A\|_2 = \rho(A)$.
On utilise aussi parfois la *norme de Frobenius* car elle est facile à calculer. Elle est définie par

$$\|A\|_F = \left(\sum_{i,j=1}^{n} |a_{ij}|^2 \right)^{1/2}.$$

Pour cette norme, on a également

$$\|AB\|_F \leq \|A\|_F \cdot \|B\|_F.$$

De plus

$$\|A\|_F^2 = tr(A^TA),$$

où *tr* désigne la trace d'une matrice, c'est-à-dire la somme des éléments de sa diagonale.

2. Conditionnement d'une matrice

Soit $A \in \mathbb{C}^{n\times n}$. Pour toute norme de Hölder, on a

$$\|AA^{-1}\| = \|I\| \leq \|A\| \cdot \|A^{-1}\|.$$

Or, la norme de la matrice identité est égale à 1 d'après la définition. Il s'en suit que l'on a l'inégalité (l'indice k est sous-entendu)

$$\kappa(A) = \|A\| \cdot \|A^{-1}\| \geq 1.$$

Ce nombre $\kappa(A)$ s'appelle le *conditionnement* de A.
On a les propriétés suivantes

1. $\kappa(\lambda A) = \kappa(A), \quad \forall \lambda \neq 0,$
2. $\kappa(A^{-1}) = \kappa(A)$ puisque A et A^{-1} jouent les rôles symétriques dans la définition de $\kappa(A)$,
3. $0 < \dfrac{1}{\|A^{-1}\|} \leq \dfrac{\|Ax\|}{\|x\|} \leq \|A\|, \quad \forall x.$

La notion de conditionnement d'une matrice est absolument fondamentale pour les méthodes de résolution des systèmes d'équations linéaires. On dit que la matrice A est *bien conditionnée* si $\kappa(A)$ est "voisin" de 1. Si $\kappa(A)$ est "grand" par rapport à 1, on dit que A est *mal conditionnée*. Naturellement les adjectifs "voisin" et "grand" sont subjectifs. Si la précision de l'ordinateur avec lequel on travaille est de 10^{-7} un conditionnement de 10^5 sera considéré comme "grand". Par contre, si la précision est de 10^{-16}, un tel conditionnement sera "petit".

2.1 Remarque.
On fait souvent l'erreur de croire qu'une matrice dont le déterminant est très voisin de zéro est mal conditionnée et cela parce qu'elle est proche d'une matrice singulière. Il n'en est rien. En effet, considérons la matrice diagonale de dimension n dont tous les termes sont égaux à un nombre ε. Son déterminant vaut ε^n alors que son conditionnement est égal à 1.

3. Préconditionnement d'un système linéaire

Soit à résoudre le système d'équations linéaires $Ax = b$. Si A est mal conditionnée, une petite perturbation des données du système linéaire (c'est-à-dire sur la matrice A et/ou le second membre b) pourra engendrer une grande variation dans la solution exacte. De plus, dans ce cas, les erreurs provenant de l'arithmétique à précision finie des ordinateurs pourront affecter les résultats d'une erreur importante. Si le conditionnement est mauvais, un certain nombre de méthodes itératives convergeront lentement car plus le conditionnement est grand et plus leur vitesse de convergence est faible. C'est le cas, par exemple, de la méthode de la plus profonde descente et de la méthode du gradient conjugué qui seront étudiées dans ce livre. Des méthodes convergentes pourront même ne plus converger du tout. Une manière de remédier à ce problème est de préconditionner le système.

Le *préconditionnement* consiste à remplacer le système $Ax = b$ par le système

$$CAx = Cb.$$

La matrice C est choisie de sorte que le conditionnement de CA soit plus petit que celui de A. Une telle stratégie s'appelle *préconditionnement* (à gauche) et la matrice C s'appelle *préconditionneur* (à gauche). Plus $\kappa(CA)$ sera voisin de 1, meilleur sera le préconditionnement. Il est évident que le meilleur préconditionneur possible est $C = A^{-1}$, un choix impossible en pratique. On prendra donc pour C une approximation de A^{-1}. Il est possible de construire de telles approximations en théorie. Cependant, leur utilisation pratique est souvent limitée aux systèmes de dimension réduite tels que C puisse être gardée en mémoire de l'ordinateur. Pour les grands systèmes, il n'existe pas de préconditionneurs valables pour toute matrice et chaque cas particulier doit être étudié individuellement.
On peut également définir un préconditionnement à droite, c'est-à-dire où l'on considère le système $ACy = b$ avec $x = Cy$ ainsi qu'un préconditionnement bilatéral $CAC'y = Cb$ avec $x = C'y$.
Il est évident que, dans ces stratégies, on ne calcule ni le produit CA ni AC'. En effet, dans les méthodes itératives, il est seulement nécessaire de savoir calculer des produits CAv où v est un vecteur quelconque. Si C est donnée, on calcule d'abord le vecteur Av puis on le multiplie par la matrice C. La même stratégie s'applique aux produits $AC'v$.

Au lieu de chercher une matrice C qui soit une bonne approximation de A^{-1}, on peut aussi chercher une matrice M qui soit une approximation de A puis prendre $C = M^{-1}$. Le système préconditionné (à gauche) s'écrit alors

$$M^{-1}Ax = M^{-1}b.$$

Dans la pratique, on ne calcule jamais la matrice M^{-1}. Pour obtenir le vecteur $u = M^{-1}Av$, on résout le système $Mu = Av$. Il faut, bien sûr, que ce système soit plus facile à résoudre que le système initial $Ax = b$. En général, on prendra pour M une matrice ayant plus d'éléments nuls que la matrice A. Par exemple, si M est la matrice diagonale formée

par la diagonale de A, M^{-1} sera une bonne approximation de A^{-1} si A est à diagonale dominante, c'est-à-dire si $\forall i, |a_{ii}| > \sum_{j \neq i} |a_{ij}|$. Les matrices M qui apparaissent dans les méthodes de relaxation (voir Section 2. du Chapitre II) peuvent aussi jouer le rôle de préconditionneur.

D'autres préconditionneurs peuvent être construits par décomposition orthogonale (voir [4]). Si $A = QU$ où Q est une matrice orthogonale (i.e. $Q^T = Q^{-1}$) et U une matrice triangulaire supérieure, alors il est possible de prendre $C = U^{-1}$. En effet, dans ce cas, $\kappa_2(AC) = \kappa_2(Q) = 1$. De même, si $A^T = QU$ et si l'on prend $C = U^{-T}$, alors $\kappa_2(CA) = \kappa_2(Q^T) = 1$. Des approximations de U^{-1} et de U^{-T} peuvent s'obtenir par décomposition orthogonale incomplète. Dans la pratique, il est nécessaire de recourir à de tels procédés incomplets pour la raison évoquée plus haut. En effet, même si les matrices A et U sont creuses (c'est-à-dire qu'une grande majorité de leurs éléments sont nuls), il n'y a aucune raison pour qu'il en soit de même pour U^{-1}. On se livre donc seulement à une décomposition incomplète qui consiste à remplacer certains éléments non nuls de U^{-1} par des zéros.
On peut également se livrer à une décomposition LU incomplète de la matrice A, une procédure connue sous le nom de ILU(0) où la première lettre signifie *incomplète*. Dans ce cas, on cherche une matrice L triangulaire inférieure à diagonale unité et une matrice U triangulaire supérieure, en imposant en plus à certains éléments de ces deux matrices d'être nuls, telles que $A = LU + R$. On prendra alors $M = LU$ comme approximation de A. On peut recommencer la décomposition LU de cette matrice M, une procédure qui s'appelle ILU(1) et qui peut être poursuivie pour obtenir des décompositions incomplètes successives ILU(k). Ces décompositions incomplètes peuvent servir également dans le préconditionnement bilatéral. En effet, si, d'après la décomposition de Gauss (voir [4]), on a $A = LU$, alors on pourra considérer le système $CAC'y = Cb$ avec $x = C'y$ et où C est une approximation de L^{-1} et C' une approximation de U^{-1}.
Lorsque, dans le préconditionnement bilatéral, les matrices C et C' sont diagonales, on parle d'*équilibrage* de la matrice A. On voit souvent écrit qu'il faut choisir C et C' de sorte que les éléments les plus grands en valeur absolue dans chaque ligne et dans chaque colonne soient à peu près égaux. C'est là une idée fausse. La bonne stratégie consiste en un choix tel que les sommes des valeurs absolues des éléments de chaque ligne et de chaque colonne de la matrice CAC' soient à peu près égales.
Le préconditionnement d'un système linéaire est une opération fondamentale, souvent même plus importante que la méthode de résolution utilisée par la suite. Malheureusement, il n'existe pas de préconditionneur universel et il est nécessaire d'en bâtir un adapté à chaque problème (ou classe de problèmes) considéré. Quand on utilise une méthode itérative pour résoudre le système $Ax = b$ on peut envisager d'utiliser une suite de préconditionneurs (C_k) ou (M_k), mais c'est une technique plus coûteuse et difficile à mettre en œuvre.
Sur le préconditionnement, voir [26] et [63].

4. Estimations de l'erreur et tests d'arrêt

Quand on utilise dans la pratique une méthode itérative pour résoudre un système linéaire, on obtient une suite de vecteurs x_k qui sont des approximations de la solution exacte x qui est uniquement obtenue à la limite. Il est, bien entendu, nécessaire d'arrêter les itérations. Il faut alors être capable de savoir si le dernier itéré calculé est une bonne approximation de la solution. C'est un problème difficile. Naturellement, si la méthode est convergente,

la suite des résidus $r_k = b - Ax_k$ va tendre vers zéro. On peut donc décider d'arrêter les itérations lorsque $\|r_k\|$ est suffisamment petit. Cependant, si la matrice A est mal conditionnée, le résidu peut être petit alors que l'erreur est grande. On peut alors penser à stopper les itérations lorsque $\|r_k\|/\|r_0\|$ est suffisamment petit, c'est-à-dire lorsque l'on a réduit suffisamment la norme du résidu initial. On peut également songer à arrêter les itérations lorsque $\|x_{k+1} - x_k\|$ est suffisamment petit. Ce genre de tests sera utilisé, en particulier, dans les méthodes itératives étudiées dans la Section 2.12 du Chapitre II.

Donnons maintenant un certain nombre de critères pour arrêter des itérations. Sauf lorsque la matrice est très bien conditionnée, aucun de ces tests n'est entièrement fiable dans tous les cas. Il vaudra donc mieux utiliser simultanément plusieurs tests et ne s'arrêter que lorsqu'ils auront tous été satisfaits. Parmi ces tests, il est également fortement conseillé d'en prévoir un sur le nombre maximum d'itérations à ne pas dépasser.

Soit x la solution exacte du système $Ax = b$ et x_k une approximation de x obtenue par une méthode itérative. L'erreur $e_k = x - x_k$ et le résidu $r_k = b - Ax_k$ sont reliés par $r_k = Ae_k$ et, donc, il n'est pas possible de calculer la norme de l'erreur à partir de celle du résidu. Il est évident que l'erreur tend vers zéro si et seulement si le résidu tend vers zéro.
On a (voir [4])

$$\frac{\|r_k\|}{\|A\|} \leq \|e_k\| \leq \|A^{-1}\| \cdot \|r_k\|. \tag{I.1}$$

Par conséquent, les quantités $\|r_k\|/\|A\|$ et $\|A^{-1}\| \cdot \|r_k\|$ peuvent être prises comme estimations de $\|e_k\|$. Cependant ces estimations demandent la connaissance de la norme de A ou de son inverse et, de plus, dans certains cas, ces bornes peuvent être beaucoup trop larges.

On a également la majoration suivante de l'erreur relative

$$\frac{\|x - x_k\|}{\|x\|} \leq \kappa(A) \frac{\|r_k\|}{\|b\|}$$

où $\kappa(A) = \|A\| \cdot \|A^{-1}\|$ est le conditionnement de la matrice A. Bien que la valeur de ce conditionnement soit, en général, inconnue on arrête souvent une méthode itérative lorsque $\|r_k\|/\|b\|$ est suffisamment petit. C'est alors à l'utilisateur de savoir si ce test d'arrêt est justifié selon la valeur de $\kappa(A)$. Cependant, dans bien des cas, ce conditionnement est totalement inconnu. Il faut donc disposer d'autres tests.

Étudions maintenant une estimation de la norme de l'erreur $\|e_k\|$ par la quantité

$$\rho_k = \frac{\|r_k\|^2}{\|Ar_k\|}.$$

Nous avons

$$\frac{\|e_k\|}{\rho_k} = \frac{\|e_k\|}{\|r_k\|^2} \|Ar_k\| = \frac{\|A^{-1}r_k\| \cdot \|Ar_k\|}{\|r_k\|^2} \leq \|A^{-1}\| \cdot \|A\| = \kappa(A).$$

D'un autre côté, $r_k = A^{-1}Ar_k$ et donc $\|r_k\| \leq \|A^{-1}\| \cdot \|Ar_k\|$. Nous avons également $\|r_k\| \leq \|A\| \cdot \|e_k\|$, d'où, en multipliant ce deux inégalités entre elles,

$$\frac{1}{\kappa(A)} \leq \frac{\|e_k\|}{\rho_k} \leq \kappa(A).$$

Ces inégalités montrent que, si le conditionnement de A est voisin de 1, alors on est certain que ρ_k est une bonne estimation de $\|e_k\|$. Cependant, même si le conditionnement de A est grand, cette estimation peut être satisfaisante.

Il est possible d'obtenir des bornes un peu plus fines, mais seulement théoriques car elles ne sont pas calculables en pratique

$$\frac{1}{\kappa(A)} \leq m(A) \leq \frac{\|e_k\|}{\rho_k} \leq M(A) \leq \kappa(A)$$

avec

$$\begin{aligned} m(A) &= \min_{u \neq 0} \frac{\|Au\| \cdot \|A^{-1}u\|}{\|u\|^2} \\ M(A) &= \max_{u \neq 0} \frac{\|Au\| \cdot \|A^{-1}u\|}{\|u\|^2}. \end{aligned}$$

Des démonstrations similaires sont valables pour la quantité $\rho'_k = \|r_k\|^2/\|A^T r_k\|$ qui est également une estimation de la norme de l'erreur. Cette estimation est, dans le cas de la norme euclidienne, une borne inférieure de la norme de l'erreur. En effet

$$\|r_k\|_2^2 = (r_k, r_k) = (r_k, Ae_k) = (A^T r_k, e_k) \leq \|A^T r_k\|_2 \cdot \|e_k\|_2.$$

Il existe des tests d'arrêt spécifiques pour certaines méthodes comme l'algorithme du gradient conjugué. Nous n'en parlerons pas ici car ils font partie d'ouvrages plus spécialisés ; voir, par exemple, [40, pp. 284-302].

Chapitre II
Méthodes itératives de base

Soit à résoudre le système d'équations linéaires réelles (pour simplifier)

$$Ax = b.$$

Une *méthode itérative* consiste à construire une suite de vecteurs qui, sous certaines conditions, converge vers la solution du système.

Dans ce Chapitre, on étudiera d'abord les méthodes de relaxation, puis la méthode des directions alternées et enfin les méthodes de Richardson. Les Chapitres suivants seront consacrés aux méthodes de projection qui, bien qu'étant des méthodes itératives, sont de nature complètement différente. C'est pour cette raison qu'elles ont été mises à part.
Pour des résultats plus complets, on pourra consulter l'ouvrage classique de Varga [51], ou les livres de Ciarlet [5], Hackbusch [31], Meurant [40] et Stewart [48].

Nous allons commencer par établir un certain nombre de résultats généraux dont nous aurons besoin dans ce Chapitre.

1. Convergence de matrices

La définition principale est

1.1 Définition.
Soit (A_k) une suite de matrices. On dit que (A_k) converge vers A si $\lim_{k\to\infty} \|A_k - A\| = 0$.
On dit qu'une matrice carrée A est convergente si la suite (A^k) de ses puissances converge vers 0 *lorsque k tend vers l'infini.*

Cherchons la condition nécessaire et suffisante pour que A soit convergente.

Il existe une matrice S régulière telle que $SAS^{-1} = J$ où J est de la *forme de Jordan*

$$J = \begin{pmatrix} J_1 & & & \\ & J_2 & & \\ & & \ddots & \\ & & & J_r \end{pmatrix},$$

J_m $(m = 1, \ldots, r)$ étant une matrice $n_m \times n_m$ de la forme

$$J_m = \begin{pmatrix} \lambda_m & 1 & & \\ & \ddots & \ddots & \\ & & \ddots & 1 \\ & & & \lambda_m \end{pmatrix},$$