Résolution des systèmes linéaires $Ax = b$. Méthodes directes de Gauss et de Householder.

Le 14 octobre 2014

Jean Roux

Mini-cours donné en 3^e année de
Licence de Sciences du département de Géosciences
École normale supérieure, Paris

Résolution des systèmes linéaires $Ax = b$, par des méthodes numériques directes.

Dans ce cours on cherche à résoudre le problème linéaire suivant : étant donnés la matrice $A \in \mathbb{R}^{n,n}$ et le vecteur $b \in \mathbb{R}^n$, trouver le vecteur $x \in \mathbb{R}^n$ solution du système

$$Ax = b. \tag{1.0.1}$$

La résolution des systèmes linéaires s'opère par deux types de méthodes : les méthodes directes et les méthodes itératives. Les méthodes directes sont principalement utilisées pour les systèmes linéaires où la matrice est pleine (peu d'éléments nuls) ou sans structure particulière (du type structure-bande, structure-bloc) et/ou propriété d'un certain type (matrice à diagonale dominante, matrice symétrique définie positive) qui apparaît fréquemment par la discretisation des équations aux dérivées partielles par des méthodes dites des différences finies ou d'éléments finis. Les systèmes linéaires avec structure ou qui possèdent peu d'éléments non nuls (systèmes creux) sont principalement résolus par les méthodes itératives.

On ne sait pas toujours a priori si la matrice A est inversible, c'est-à-dire si l'équation (1.0.1) a une solution unique. Il est donc souhaitable de disposer de méthodes qui en décident. Les méthodes directes exposées ci-après le permettent.

Nous allons examiner les méthodes de Gauss et de Householder. La méthode de Gauss peut s'utiliser lorsque la matrice A est constituée d'éléments complexes ($a_{i,j} \in \mathbb{C}$). Résoudre le système linéaire complexe $Az = b$ où $b \in \mathbb{C}^n$ et $A \in \mathbb{C}^{n,n}$ revient à trouver $z \in \mathbb{C}^n$ sous la forme $z = x + iy$ où $(x, y) \in \mathbb{R}^n \times \mathbb{R}^n$. En décomposant la matrice sous la forme $A = A_1 + iA_2$ où $A_1 \in \mathbb{R}^{n,n}$ et $A_2 \in \mathbb{R}^{n,n}$ et le second membre $b \in \mathbb{C}^n$ sous la forme $b = b_1 + ib_2$ où $(b_1, b_2) \in \mathbb{R}^n \times \mathbb{R}^n$, on est conduit à résoudre un système linéaire réel d'ordre $2n$ en les inconnues x et y dont l'écriture est évidente. On peut aussi travailler sur l'algorithme de Gauss en considérant les opérations algébriques directement sur le corps des complexes, naturellement plus coûteuses que celles faites sur le corps des réels.

Par ailleurs la méthode de Gauss peut résoudre des systèmes linéaires $Ax = b$ où la matrice A est rectangulaire, à savoir : trouver $x \in \mathbb{R}^p$ tel que $Ax = b$ où $A \in \mathbb{R}^{n,p}$ et $b \in \mathbb{R}^n$. Mais alors l'analyse matricielle de la méthode ne fournit pas la "fameuse" décomposition LU de la matrice A, où L est une matrice triangulaire inférieure avec des 1 sur la diagonale et U est une matrice triangulaire supérieure. Cette décomposition LU étant indispensable, par exemple, à la mise en oeuvre de la méthode "LR" de Rutishauser de calcul des valeurs propres

Aussi nous nous limiterons aux systèmes carrés et réels avec $A \in \mathbb{R}^{n,n}$ et $b \in \mathbb{R}^n$. En pratique l'entier n peut être très grand ; pour certains "gros" problèmes le nombre n est de l'ordre du million, voire plus.

Précisons enfin que la résolution d'un problème de ce type intervient dans de très nombreux problèmes de l'analyse numérique. L'étude des méthodes de résolution est donc indispensable.

1.1 Conditionnement et stabilité

Un souci constant en analyse numérique doit être la précision des valeurs calculées. Tous les calculs étant effectués avec des erreurs d'arrondi, l'effet de ces erreurs peut être négligeable ou catastrophique. Mais aussi, le problème à résoudre peut présenter intrinsèquement un caractère d'*instabilité*. Considérons par exemple le système :

$$\begin{cases} 2x + 6y = 8 \\ 2x + 6.00001y = 8.00001 \end{cases}.$$

En "portant" la première équation dans la seconde, on a $8-6y+6.00001y = 8.00001$, soit $0.00001y = 0.00001$ d'où $y = 1$ et $x = 1$.

Soit maintenant le système "voisin" :

$$\begin{cases} 2x + 6y = 8 \\ 2x + 5.99999y = 8.00002 \end{cases}$$

De même, on a maintenant $8-6y+5.99999y = 8.00002$, soit $-0.00001y = 0.00002$, d'où $y = -2$ et $x = 10$, qui est loin d'être "voisine" de la solution précédente. On dit d'un tel système qu'il est *mal conditionné*. Notons que ce conditionnement est défini par le produit des normes matricielles $\|A\|\|A^{-1}\|$ pour une norme matricielle (subordonnée) quelconque, ce qui exige de connaître l'inverse A^{-1} de A. L'étude du *conditionnement* n'est pas abordée dans ce cours (voir livre).

1.2 Méthodes numériques directes

Pour les méthodes directes trois idées sont à retenir.

- On ne calcule jamais A^{-1}.
- On cherche à transformer la matrice (évidemment sans changer la solution du système) en une forme triangulaire supérieure qui est facilement résoluble par une récurrence de "bas en haut". En effet, on calcule alors directement la dernière composante de la solution par la dernière ligne de la matrice, puis l'avant-dernière composante de la solution par l'avant-dernière ligne de la matrice et, de proche en proche, la première composante de la solution par la première ligne de la matrice, les $(n-1)$ dernières composantes étant calculées.

- Il est facile de compter le nombre d'opérations élémentaires (additions et multiplications) que l'on utilise. Ces méthodes sont dites *directes* car elles calculent la solution en un nombre *fini* (qui peut être grand) et connu d'opérations. Il n'y a pas de test d'arrêt de la méthode, on calcule directement la solution. Dès lors on évalue les performances d'une méthode directe en comptant les opérations élémentaires ; on calcule ainsi ce que l'on appelle la *complexité* de la méthode. Naturellement il est alors possible (puisque l'on travaille en arithmétique finie) d'étudier la sensibilité de la solution par rapport aux données, sensibilité due à la propagation aux erreurs d'arrondis.

Remarque 1.2.1. *Pourquoi ne pas utiliser les formules de Cramer ? On rappelle que* $x_k = \det(A_k)/\det(A)$*, où* A_k *est la matrice carrée obtenue en remplaçant la* k^{eme} *colonne de* A *par le le vecteur* b *de (1.0.1). Ces formules reviennent à calculer* $(n+1)$ *déterminants. Or un déterminant exige* $(n-1)n!$ *multiplications et* $(n!-1)$ *additions. L'algorithme serait donc non polynomial du point de vue du nombre des opérations élémentaires (op.él.) et, par conséquent, totalement inutile dès que* n *dépasse la cinquantaine. Par exemple si* $n = 10$ *cela donne environ 400 millions d'op.él.. En comparaison, nous verrons (voir Remarque 1.2.5) que la méthode de Gauss demande environ* $2n^3/3$ *op.él. (cette estimation est très grossière, mais elle précise l'ordre de grandeur), ce qui donne seulement environ 670 op.él. pour* $n = 10$*.*

1.2.1 Méthode de Gauss : description pratique

On suppose, pour présenter la méthode, que la matrice A est *a priori* régulière. On verra (voir Remarque 1.2.3) que la méthode de Gauss statut sur la régularité de la matrice A. On pose $A = A^{(1)} = (a_{ij}^{(1)})$. On suppose donc que $\det(A^{(1)}) \neq 0$. On souhaite que $a_{11}^{(1)} \neq 0$. Si tel n'est pas le cas, on prémultiplie A par une matrice de permutation élémentaire $P_{\lambda,\mu}$ *ad-hoc*, c'est le *pivotage*. L'introduction (éventuelle) de la matrice $P_{\lambda,\mu}$ n'a rien à voir avec la régularité de A. Soit donc

$$\mathcal{A}^{(1)} = P_{1,\zeta_1} A^{(1)} = (\alpha_{ij}^{(1)}). \tag{1.2.1}$$

avec $\alpha_{11}^{(1)} \neq 0$. Dans la méthode de Gauss on travaille toujours avec des matrices de permutation élémentaire, le mot élémentaire sera généralement omis dans ce qui suit. La prémultiplication de A par la matrice de permutation P_{1,ζ_1} permute les lignes 1 et ζ_1 de la matrice A ; comme la matrice A est supposée régulière la première colonne n'est pas nulle et il est toujours possible de permuter, si $a_{11}^{(1)} = 0$, la ligne 1 et une ligne ζ_1 caractérisée par $a_{\zeta_1,1}^{(1)} \neq 0$. Si $a_{11}^{(1)} \neq 0$ le pivotage ne semble pas nécessaire *a priori* et l'on peut croire qu'il suffit de poser $P_{1,\zeta_1} = I$. Mais sur ordinateur on travaille toujours en arithmétique non exacte, c'est-à-dire en virgule flottante,

et s'il existe plusieurs éléments non nuls dans la première colonne on verra que, pour des raisons d'erreurs d'arrondis, il est nécessaire de choisir le plus grand élément en module des éléments non nuls, ce qui nécessite presque toujours de choisir $P_{1,\zeta_1} \neq I$. Dans le cas assez rare (ce qui justifie notre locution "presque toujours") où $a_{11}^{(1)}$ est le plus grand élément en module des éléments de la première colonne, il n'est pas nécessaire de faire une permutation des lignes 1 et ζ_1, ce qui revient matriciellement à faire $P_{1,\zeta_1} = P_{1,1} = I$ dont le déterminant vaut 1.

Maintenant commence la triangulation. On s'efforce d'annuler tous les termes de la première colonne de $\mathcal{A}^{(1)}$ à l'exception du premier. Soit donc

$$A^{(2)} = E^{(1)}\mathcal{A}^{(1)}$$

avec

$$E^{(1)} = \begin{vmatrix} 1 & & & & \\ -\frac{\alpha_{21}^{(1)}}{\alpha_{11}^{(1)}} & 1 & & & \\ \vdots & 0 & \ddots & 0 & \\ \vdots & & & \ddots & \\ -\frac{\alpha_{n1}^{(1)}}{\alpha_{11}^{(1)}} & 0 & \dots & & 1 \end{vmatrix}.$$

La matrice $E^{(1)}$ est construite pour annuler tous les termes de la première colonne à l'exception du premier terme. L'intérêt de l'opération est évident en voyant (Figure 1.1) la structure de la matrice $A^{(2)}$ obtenue.

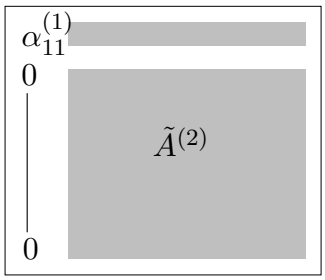
$A^{(2)} =$ 

Figure 1.1: Premiere etape de la methode de Gauss.

La triangulation est amorcée et le processus va pouvoir se poursuivre. En effet $A^{(2)} = E^{(1)}\mathcal{A}^{(1)} = E^{(1)}P_{1,\zeta_1}A^{(1)}$ avec $\det(E^{(1)}) = 1$ car la matrice $E^{(1)}$ est triangulaire inférieure avec des 1 sur la diagonale. On rappelle de plus, que si P est une matrice de permutation élémentaire alors $\det(P) = -1$ (voir livre). On a donc $\det(A^{(2)}) = \pm\det(A^{(1)})$ (le signe $+$ prenant en compte l'éventualité d'une non permutation avec $P_{1,\zeta_1} = I$) et la matrice $A^{(2)}$ est régulière car son déterminant est non nul. Comme $\det(A^{(2)}) = \alpha_{11}^{(1)}\det(\tilde{A}^{(2)})$ (Figure 1.1) et comme, par construction, $\alpha_{11}^{(1)} \neq 0$ on en déduit que $\det(\tilde{A}^{(2)}) \neq 0$. La matrice $\tilde{A}^{(2)}$ est régulière, sa première

colonne est non nulle, et on peut appliquer à cette matrice l'algorithme précédent. L'algorithme de triangulation peut se poursuivre avec prémultiplication, à chaque étape, d'une matrice de permutation puis par une matrice $E^{(k)}$ dont on comprend aisèment la structure (voir Figure 1.2) à partir de celle de $E^{(1)}$.

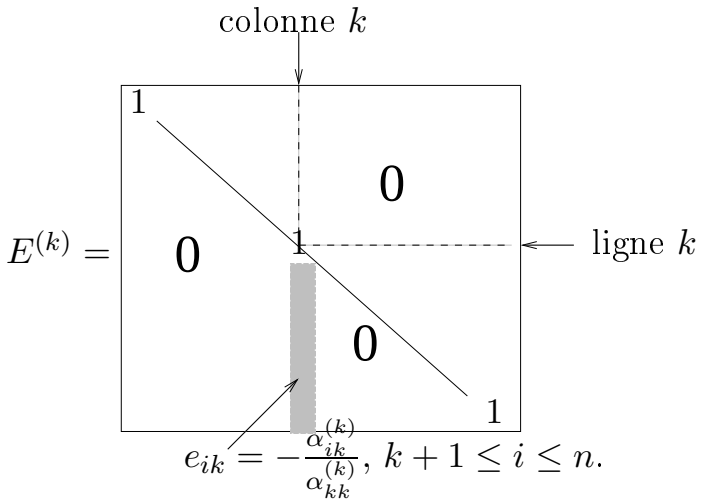


Figure 1.2: Structure de E_k.

Évidemment il ne faut pas oublier, à chaque étape, de permuter aussi les composantes du second membre et de les multiplier par les $E^{(i)}$. À la fin de l'algorithme (en n étapes) on obtient un système du type

$$Ux = b^{(n)},$$

où U est une matrice triangulaire supérieure. Sa résolution par récurrence est évidente comme nous l'avons déjà examiné.

Mais l'algorithme ne peut pas s'appliquer tel quel : une stratègie de choix de pivot (on appelle pivot l'élèment non nul choisi dans la colonne, c'est l'élément $\alpha_{11}^{(1)}$ à la première étape) est nécessaire. L'exemple suivant va nous en convaincre.

Travaillons avec un calculateur à trois chiffres significatifs (en virgule flottante), la mantisse est un nombre compris entre 0 et 10 et on ne peut additionner des nombres que s'ils ont même exposant. Cherchons à résoudre le système

$$\begin{cases} 10^{-4}x + y = 1 \\ \quad x + y = 2 \end{cases}.$$

La solution exacte est

$$\begin{cases} x = 1,00010 \fallingdotseq 1 \\ y = 0,99990 \fallingdotseq 1 \end{cases},$$

où le symbole $\fallingdotseq$ désigne la valeur réellement représentée dans le calculateur.

Que fait l'algorithme sans stratégie de pivot ?

Comme $a_{11}^{(1)} = 10^{-4} \neq 0$ nous pouvons commencer *a priori* le calcul avec confiance, sans prémultiplication par une matrice de permutation (ce qui revient à prendre $P = I$). Après multiplication par la matrice

$$E^{(1)} = \begin{vmatrix} 1 & 0 \\ -10^4 & 1 \end{vmatrix},$$

le système initial devient le système triangulaire supérieure suivant :

$$\begin{cases} 10^{-4}x + y = 1 \\ (-10^4 + 1)y = -10^4 + 2 \end{cases}.$$

Traduisons les chiffres en virgule flottante avec trois chiffres significatifs

$$\begin{cases} -10^4 + 1 = -1,000 \times 10^4 + 0,0001 \times 10^4 \fallingdotseq -1,000 \times 10^4 \\ -10^4 + 2 = -1,000 \times 10^4 + 0,0002 \times 10^4 \fallingdotseq -1,000 \times 10^4 \end{cases}.$$

La deuxième équation donne alors immédiatement $y \fallingdotseq 1$, ce qui est correct. Mais en reportant cette valeur dans la première équation on trouve $x \fallingdotseq 0$, ce qui est tout à fait irréaliste !

Essayons alors, en échangeant la première ligne avec la deuxième, de pivoter avec $a_{21}^{(1)} = 1$, la matrice $E^{(1)}$ s'écrit

$$E^{(1)} = \begin{vmatrix} 1 & 0 \\ -10^{-4} & 1 \end{vmatrix},$$

et le système devient (n'oublions pas la permutation préalable)

$$\begin{cases} x + y = 2 \\ (-10^{-4} + 1)y = -2.10^{-4} + 1 \end{cases},$$

avec les représentations suivantes des coefficients : $1 \fallingdotseq 1,000 \times 10^0$ et $-10^{-4} \fallingdotseq -0,0001 \times 10^0 \fallingdotseq 0$, le système est "traduit" en machine par le système

$$\begin{cases} x + y \fallingdotseq 2 \\ y \fallingdotseq 1 \end{cases},$$

il vient donc $y \fallingdotseq 1$ d'où, en reportant dans la première équation $x \fallingdotseq 1$, l'algorithme est sauf !

Cet exemple montre clairement la nécessité d'une stratégie de pivotage. Deux techniques sont utilisées en pratique

- pivot partiel : on prend le plus grand élément sous diagonal (terme diagonal inclus), en module, de la colonne dont on cherche à annuler les termes sous-diagonaux.

- pivot total : on prend le plus grand élément, en module, de toute la sous-matrice $\tilde{A}^{(2)}$ (resp. $\tilde{A}^{(k)}$) à l'étape 2 (resp. k) et on utilise les permutations convenables (des lignes par prémultiplication par une matrice du type P, des colonnes par postmultiplication par une matrice de permutation du même type).

Remarque 1.2.2. *En pratique la première technique (un peu moins coûteuse en temps calcul) est très souvent suffisante.*

Remarque 1.2.3. *A priori on ne sait pas toujours si la matrice A est régulière. Observons la configuration de la triangularisation à l'étape k (Figure 1.3). Il est clair que* $\det(A^{(k+1)}) = \left(\prod_{i=1}^{k} \alpha_{ii}^{(i)}\right) \det(\tilde{A}^{(k+1)})$ *et que, de*

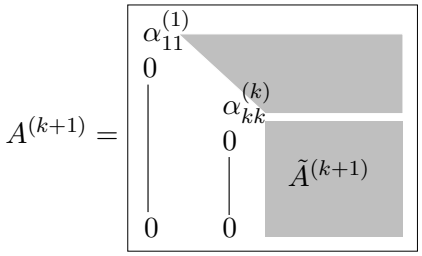


Figure 1.3: k-ième étape de la méthode de Gauss.

plus, $\det(A^{(1)}) = (-1)^l \det(A^{(k+1)})$ *l'entier l étant le nombre de permutations de lignes (i.e. le nombre de pivotages) effectuées jusqu'à l'étape k de l'algorithme. Ainsi, si la matrice $A(= A^{(1)})$ est singulière, à une certaine étape k la matrice $\tilde{A}^{(k+1)}$ sera singulière, cette singularité apparaissant avec une première colonne constituée d'éléments nuls (en pratique le test de nullité du pivot est assez délicat à choisir). L'algorithme de Gauss permet donc de décider de la singularité d'une matrice.*

Remarque 1.2.4. *Dans le cas où A est régulière on voit que* $\det(A) = (-1)^p \det(A^{(n)})$, *la matrice $A^{(n)}$ étant triangulaire supérieure ;* $\det(A^{(n)})$ *est égal au produit de ses éléments diagonaux. C'est-à-dire que son déterminant est égal, au signe près, au produit des pivots. Le signe est donné par $(-1)^p$, l'entier p étant le nombre de permutations de lignes effectuées au cours de toute la triangulation, nombre que l'on connait par la mise en œuvre de la méthode. C'est un résultat parfois utile, un peu imprévu, de la mise en œuvre de la méthode de Gauss. Ce procédé de calcul du déterminant est beaucoup moins coûteux que celui des méthodes usuelles de son évaluation (voir la Remarque 1.2.1).*

1.2.2 Analyse matricielle de la méthode de Gauss

Cas simplifié On suppose dans ce paragraphe que la matrice A est régulière et que le pivot se trouve toujours d'emblée sur la diagonale. Écrivons les for-

mules générales de la multiplication de $A^{(k)}$ par la matrice $E^{(k)}$, donnant les termes de la matrice $\tilde{A}^{(k+1)}$ (Figure 1.3).

$$a_{ij}^{(k+1)} = a_{ij}^{(k)} + e_{ik} a_{kj}^{(k)}, \quad k+1 \leq i,j \leq n, \quad \text{avec} \quad e_{ik} = -\frac{a_{ik}^{(k)}}{a_{kk}^{(k)}}, \quad k+1 \leq i \leq n. \tag{1.2.2}$$

Notons que les k premières lignes de $A^{(k)}$ sont conservées et que les termes de la $(k)^e$-colonne de $A^{(k+1)}$, au-dessous de $a_{kk}^{(k)}$, sont nuls (Figure 1.3).

Remarque 1.2.5. *Essentiellement ce sont les formules (1.2.2) qui donnent le nombre des op.él. de la méthode de Gauss. En effet la résolution du système $Ux = b^{(n)}$ où U est une matrice triangulaire supérieure se fait approximativement en n^2 op.él. (un ordre de grandeur de moins que l'algorithme de triangulation proprement dit). Notons que les permutations d'éléments ne sont pas comptabilisées comme op.él. bien que, naturellement, dans la recherche du coût, il faut prendre en compte le temps de recherche des pivots. On voit sur les formules (1.2.2) que l'on fait, à chaque appel, une addition et une multiplication. Un comptage munitieux montre qu'il faut, en négligeant les termes en n^2 et en n (ce qui est légitime dès que n est assez grand), $n^3/3$ additions et $n^3/3$ multiplications pour la mise en œuvre de Gauss. Ce calcul utilise la formule $\sum_{k=1}^{n} k^2 = n(n+1)(2n+1)/6$. Notons enfin que la méthode nécessite a priori de l'ordre de $n^2/2$ divisions dont le temps d'exécution est supérieur à la multiplication ; aussi dans le calcul de e_{ik} (voir (1.2.2)) il est donc très recommandé de calculer d'abord l'inverse de $a_{kk}^{(k)}$ (ce qui nécessite au total seulement n divisions) et ensuite de calculer les e_{ik} pour $i \geq (k+1)$ par multiplication.*

Inversement le passage de l'itération $(k+1)$ à l'itération (k) s'écrit donc :

$$a_{ij}^{(k)} = a_{ij}^{(k+1)} - e_{ik} a_{kj}^{(k+1)}, \quad k+1 \leq i,j \leq n, \quad \text{car} \quad a_{kj}^{(k+1)} = a_{kj}^{(k)}.$$

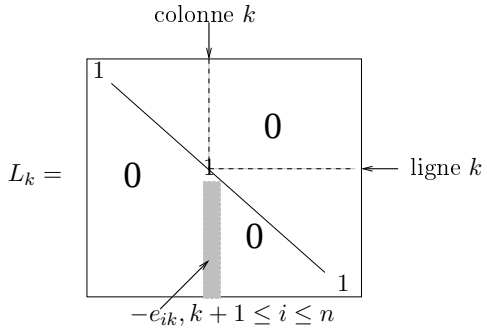


Figure 1.4: Description de L_k avec $-e_{kk} = 1$.

Soit matriciellement $A^{(k)} = L_k A^{(k+1)}$, avec L_k exhibée à la figure 1.4. De proche en proche $A^{(1)} = L_1 L_2 \cdots L_{n-1} A^{(n)}$. On observe que le produit $L_1 L_2$ s'écrit comme à la Figure 1.5. La première (resp. deuxième) colonne du produit est constituée de la colonne de L_1 (resp. L_2).

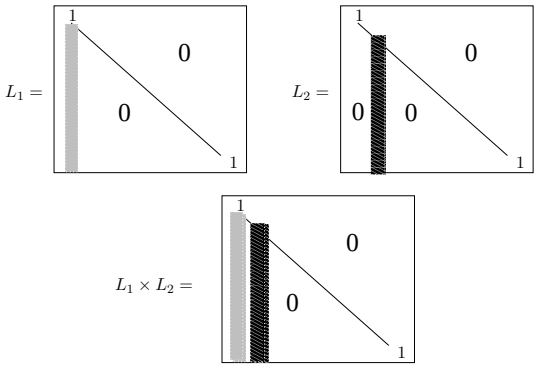


Figure 1.5:

Remarque 1.2.6. *Attention cela n'est vrai que dans cet ordre ($L_2 \times L_1$ reste triangulaire inférieure, mais certains éléments de ses deux premières colonnes sont des produits des éléments de L_1 et L_2).*

On se convaincra maintenant aisément que le produit $L = \prod_{i=1}^{n-1} L_i$ est triangulaire inférieure (avec des 1 sur la diagonale). Finalement on a

$$A = A^{(1)} = LA^{(n)} = LU,$$

où U est triangulaire supérieure. C'est, ce que l'on appelle, la décomposition "LU" de la matrice A.

On peut donc effectuer facilement (étant donné les structures des matrices) les calculs suivants :

$$\begin{cases} Ly = b \\ Ux = y \end{cases}.$$

Remarque 1.2.7. *En pratique, dans le cas d'un seul système à calculer, on résout $Ux = b^{(n)}$, avec $b^{(n)}$ calculé en cours de triangularisation. Mais si l'on doit résoudre plusieurs systèmes avec la même matrice A mais avec des seconds membres différents, il est évidemment très utile de garder en mémoire la décomposition $A = LU$ qui n'est faite qu'une seule fois, et de résoudre comme indiqué les systèmes triangulaires précédents pour les différents seconds membres.*

Cas du pivot partiel On montre (voir livre) que la stratégie de Gauss avec pivot partiel revient à trouver la décomposition "LU" d'une certaine matrice PA où P est une matrice de permutation de lignes. Finalement on

a donc à résoudre $Ax = b$, soit $PAx = Pb$ avec $PA = LU$. On effectue donc, de façon immédiate, les calculs suivants :

$$\begin{cases} Ly = Pb \\ Ux = y \end{cases} .$$

Notons enfin que le prix, en temps machine, de la recherche indispensable du pivot à chaque étape peut-être relativement coûteux. Bien que le nombre requis d'opérations élémentaires soit seulement de l'ordre de $2n^3/3$, ce surcoût altère la performance de la méthode de Gauss si elle est mesurée seulement en nombre d'op. él..

Cas du pivot total On montre (voir livre) que la stratégie de Gauss avec pivot total revient à trouver la décomposition "LU" d'une certaine matrice PAQ où P et Q sont des matrices de permutation.

Calcul de l'inverse d'une matrice Soit la décomposition $A = LU$ de la matrice A. Considérons les n systèmes linéaires $Au_j = e_j$, $1 \leq j \leq n$, où e_j est le jème vecteur de la base canonique de $\mathbb{R}^n$. Pour j fixé, la solution u_j représente le jème vecteur colonne de la matrice A^{-1}. Chacun des systèmes $Au_j = e_j$ est très rapide à résoudre si la décomposition $A = LU$ est connue. On peut donc dire que le calcul de A^{-1} est facile si la décomposition $A = LU$ de la matrice A est connue au préalable.

Terminons l'étude de la méthode de Gauss par un résultat, qui reprend en partie la remarque 1.2.3, qui affirme que l'on peut toujours trianguläriser une matrice carrée, qu'elle soit régulière ou non, par la méthode de Gauss.

Théorème 1.2.1. *Soit A une matrice carrée, inversible ou non. Il existe (au moins) une matrice inversible M telle que la matrice MA soit triangulaire supérieure.*

Preuve : On se place dans le cadre de la méthode de Gauss avec pivot partiel. Ce résultat est déjà démontré lorsque la matrice A est inversible. Supposons que la matrice A soit singulière. Par ce qui précède on sait que la matrice est singulière si et seulement si, à une étape k de la triangularisation, les éléments $\alpha_{ik}^{(k)}$, $k \leq i \leq n$, sont tous nuls (voir figure 1.3). Mais alors $\alpha_{kk}^{(k)} = 0$ et $A^{(k+1)}$ est déjà de la forme $A^{(k+2)}$; il suffit donc de considérer $P^{(k+1)} = E^{(k+1)} = I$ pour continuer l'algorithme de Gauss (sachant même que A est singulière), pour finalement trouver une matrice M inversible (avec des 1 sur la diagonale lorsque $\alpha_{kk}^{(k)} = 0$) telle que MA soit triangulaire supérieure. □

Ainsi, dans le cas où A est singulière, la matrice M, comme la matrice L de la décomposition LU obtenue dans le cas où A est régulière, a toujours un déterminant égal à 1. Cependant la matrice triangulaire supérieure MA

obtenue a alors un certain nombre m de zéros sur sa diagonale, qui peuvent être considérés comme des "pivots" nuls (par extension du mot pivot, puisqu'il désigne essentiellement un nombre non nul).

Remarque 1.2.8. *On démontre que le* rang *d'une matrice A (c'est-à-dire le rang de ses vecteurs colonnes) ne change pas par les opérations élémentaires suivantes :*

- *permutation de deux lignes,*
- *multiplication d'une ligne par un scalaire non nul,*
- *addition à une ligne d'un multiple d'une autre ligne.*

Ces opérations élémentaires sont celles de la méthode de Gauss. Ce qui veut dire que la matrice triangulaire supérieure U obtenue par la méthode de Gauss a même rang que la matrice A. Or le rang de U est égal à n-(nombre de zéros sur sa diagonale), autrement dit le rang de la matrice A est égal à n moins le nombre de "pivots" nuls dans la mise en oeuvre de la méthode de Gauss. Ce nombre de "pivots" nuls pouvant se calculer (voir la preuve du théorème 1.2.1). C'est la méthode de calcul du rang d'une matrice par la méthode de Gauss.

1.2.3 Méthode de Householder

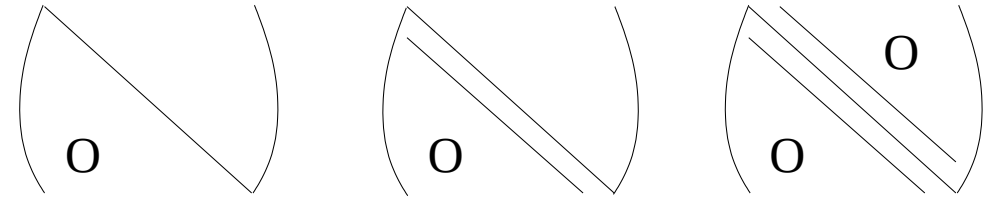
Pourquoi étudier cette méthode ? Indépendamment de son intérêt théorique, elle possède des qualités intrinsèques, mais aussi elle est indispensable, à titre d'illustration, à la mise en oeuvre de la méthode "QR" de calcul des valeurs propres d'une matrice quelconque, non symétrique par exemple. Cette méthode "QR" est la plus universellement utilisée, ce qui justifie au moins l'intérêt que l'on porte à la méthode de Householder qui construit une factorisation QU de la matrice A. Notons encore que la factorisation QU est préalable au calcul des valeurs propres d'une matrice symétrique par la méthode de "Givens-Householder".

On suppose toujours que la matrice A est réelle. Par principe la méthode de Householder ne peut travailler que sur les matrices réelles et carrées.

On va utiliser une technique générale qui permer de passer d'une matrice A sans aucune particularité à une matrice du type de la Figure 1.6.

Une matrice de Hessenberg supérieure (resp. inférieure) est une matrice triangulaire supérieure (resp. inférieure) avec la première sous-(resp. sur-) diagonale non nulle. C'est à partir des deux dernières formes que l'on peut calculer les valeurs propres de la matrice A. Ces mises en forme sont préalables au calcul des valeurs propres de A, en précisant que pour la forme tridiagonale cela est vrai spécifiquement lorsque A est symétrique. À partir de la première forme on peut encore résoudre le système $Ax = b$.

Dans ce qui suit c'est essentiellement la construction, à partir de la matrice A, de la matrice triangulaire supérieure qui va nous intéresser.



Triangulaire supérieure | Hessenberg supérieure | Tridiagonale

Figure 1.6:

Principe de la méthode On travaille avec des matrices élémentaires H dites de Householder (du nom de son inventeur) du type $H = I - 2uu^T$, $u \in \mathbb{R}^n$, $u \neq 0$ avec $u^T u = 1$ soit $\|u\|_2 = 1$. En pré-multipliant la matrice A par des matrices successives de type H, on cherche à transformer le système en un système triangulaire supérieur qui lui est équivalent.

Remarque 1.2.9. *On vérifie immédiatement que $Hu = u - 2u = -u$, ce qui signifie que -1 est valeur propre simple de H associée au vecteur propre u.*

Lemme 1.2.1. *La matrice H représente une symétrie par rapport à l'hyperplan $\mathcal{P}$ orthogonal à u défini par $\mathcal{P} = \{v; v^T u = 0\}$. De plus on a $\det(H) = -1$.*

Preuve : On a, à la fois, $Hu = -u$ et, si $v \in \mathcal{P}$, $Hv = v - 2u(u^T v) = v$. Donc tout vecteur de l'hyperplan $\mathcal{P}$ est conservé par H et donc 1 est une valeur propre de H de multiplicité $(n-1)$ associé à v, alors que le vecteur u orthogonal à $\mathcal{P}$ est transformé par H en son symétrique, à savoir que u est un vecteur propre de H associé à la valeur propre -1. L'assertion du lemme est donc démontrée. De surcroît, ce qui précède démontre que $\det(H) = (-1)(1)^{n-1} = -1$. □

Propriétés de H

- $H^T = H$, en effet on a $H^T = I - (2uu^T)^T = H$, la matrice H est symétrique.
- $HH^T = I$, là encore la vérification est immédiate. On a $HH^T = H^2 = (I - 2uu^T)(I - 2uu^T) = I - 4uu^T + 4u(u^T u)u^T = I$ car $u^T u = 1$ par hypothèse.

Il en résulte que H est une matrice (dite élémentaire) *orthogonale*. On en déduit que $H^{-1} = H^T = H$, les matrices inverses de H sont aussi des matrices de Householder élémentaires.

Construction de la matrice H

Lemme 1.2.2. *(lemme-clef de la méthode) Soit $a \in \mathbb{R}^n$ un vecteur non nul, il existe une matrice élémentaire orthogonale H et un nombre réel non nul α tels que $Ha = \alpha e_1$, où e_1 est le premier vecteur de la base canonique de $\mathbb{R}^n$.*

Preuve : Supposons le problème résolu. Il existe donc un vecteur $u \in \mathbb{R}^n$ et donc une matrice orthogonale $H = I - 2uu^T$ tels que $\|Ha\|_2 = \|a\|_2$ (en effet, si on note par $(.,.)$ désigne le produit scalaire dans $\mathbb{R}^n$, on a $\|Ha\|_2^2 = (Ha, Ha) = (H^T Ha, a) = \|a\|_2^2$ car H est orthogonale). Dès lors si le problème est résolu on a nécessairement $\|Ha\|_2 = \|\alpha e_1\|_2$ avec $\|Ha\|_2 = \|a\|_2$, comme $\|e_1\|_2 = 1$ il vient donc :

$$|\alpha| = \|a\|_2. \tag{1.2.3}$$

Comme $Ha = \alpha e_1$ il vient $(I - 2uu^T)a = \alpha e_1$ soit, en posant $\mu = 2u^T a$, $-\mu u = \alpha e_1 - a$ et donc

$$\mu u = a - \alpha e_1. \tag{1.2.4}$$

Prémultiplions enfin (1.2.4) par $2a^T$, en tenant compte de (1.2.3) il vient

$$2a^T \mu u = \mu^2 = 2\alpha^2 - 2\alpha(a^T e_1).$$

En notant a_1 la première composante du vecteur a on a donc

$$\mu^2 = 2\alpha(\alpha - a_1). \tag{1.2.5}$$

Alors, le vecteur a étant donné, (1.2.3) donne α au signe près : $\alpha = \pm\|a\|_2$. Ensuite (1.2.5) donne le scalaire μ (avec encore une indétermination du signe). Enfin (1.2.4) défini le vecteur u.

Vérifions que le vecteur u et le matrice H ainsi construits ont les propriétés requises. On suppose que $\mu \neq 0$, ce qui revient à supposer, le vecteur a étant non nul par hypothèse, que le vecteur a possède une composante non nulle autre que a_1 (sinon le problème est résolu et $H = I$). Alors, compte tenu de (1.2.4) et (1.2.5)

$$\|u\|_2^2 = \frac{(\mu u)^T(\mu u)}{\mu^2} = \frac{(a - \alpha e_1)^T(a - \alpha e_1)}{2\alpha(\alpha - a_1)},$$

ou encore d'après (1.2.3)

$$\|u\|_2^2 = \frac{\alpha^2 - 2\alpha a_1 + \alpha^2}{2\alpha(\alpha - a_1)} = \frac{2\alpha(\alpha - a_1)}{2\alpha(\alpha - a_1)} = 1,$$

ce qui implique que $\|u\|_2 = 1$, le vecteur u a la norme requise.

D'autre part, par la définition de μ, on a $Ha = (I - 2uu^T)a = a - u\mu$ et donc, d'après (1.2.4), $Ha = \alpha e_1$, la matrice H ainsi construite convient. □

L'algorithme de la construction de H utilise successivement deux extractions de racine carrée, une pour α, l'autre pour μ. Cet algorithme peut se simplifier en posant $\nu = \mu^2/2$ et $v = \mu u$. Alors, en supposant toujours $\mu \neq 0$, on a $H = I - vv^T/\nu$ et l'algorithme devient

$$\left\{ \begin{array}{l} |\alpha| = \left(\sum_{i=1}^{n} a_i^2 \right)^{\frac{1}{2}} \\ \nu = \alpha(\alpha - a_1) \\ v = a - \alpha e_1 \end{array} \right. \tag{1.2.6}$$

où on a maintenant à faire une seule extraction de racine carrée. L'ordre de calcul du vecteur v et du scalaire ν est indifférent.

Il faut lever l'indétermination sur le signe de α. Pour minimiser les erreurs d'arrondi, il faut que le ν (qui intervient au dénominateur) soit le plus grand possible en module, ce qui entraîne que α doit être du signe de $-a_1$. On prend donc

$$\alpha = -\mathrm{sgn}(a_1) \left(\sum_{i=1}^{n} a_i^2 \right)^{\frac{1}{2}}. \tag{1.2.7}$$

ce qui implique d'après (1.2.6)

$$\nu = -\alpha \left(a_1 + \mathrm{sgn}(a_1) \left(\sum_{i=1}^{n} a_i^2 \right)^{\frac{1}{2}} \right). \tag{1.2.8}$$

En résumé, a_1 étant, rappelons-le, la première composante du vecteur a, le calcul de la matrice $H = I - vv^T/\nu$ s'effectue par l'algorithme suivant :

$$\left\{ \begin{array}{l} \alpha = -\mathrm{sgn}(a_1) \left(\sum_{i=1}^{n} a_i^2 \right)^{\frac{1}{2}} \\ \nu = -\alpha \left(a_1 + \mathrm{sgn}(a_1) \left(\sum_{i=1}^{n} a_i^2 \right)^{\frac{1}{2}} \right) \\ v = a - \alpha e_1 \end{array} \right. . \tag{1.2.9}$$

Mise en œuvre de la méthode de Householder Il est nécessaire d'avoir un algorithme de calcul du produit d'un vecteur d quelconque par la matrice H. On pose $c = Hd = (I - vv^T/\nu)d$, alors $c = d - \beta v/\nu$ avec $\beta = v^T d = \sum_{i=1}^{n} v_i d_i$, d'où finalement $c = d - \gamma v$ avec $\gamma = \beta/\nu$.

En résumé, soit la matrice de Householder $H = I - vv^T/\nu$, le calcul du produit d'un vecteur d par la matrice H se fait simplement par l'algorithme suivant

$$\left\{ \begin{array}{l} \beta = v^T d = \sum_{i=1}^{n} v_i d_i \\ \gamma = \beta/\nu \\ c = Hd = d - \gamma v \end{array} \right. \tag{1.2.10}$$

Algorithme de Householder Soit A une matrice *régulière* que l'on note $A^{(1)}$ avec la notation suivante

$$A^{(1)} = \begin{pmatrix} a_{11}^{(1)} & \cdots & a_{1n}^{(1)} \\ \vdots & & \vdots \\ a_{n1}^{(1)} & \cdots & a_{nn}^{(1)} \end{pmatrix} \tag{1.2.11}$$

on note a le premier vecteur colonne de la matrice $A^{(1)}$.

Précisons seulement la première étape de l'algorithme. On construit la matrice $H^{(1)} = I - vv^T/\nu$ telle que $H^{(1)}a = \alpha e_1$ par l'algorithme (1.2.9). Alors on a

$$A^{(2)} = H^{(1)}A^{(1)} = \begin{pmatrix} a_{11}^{(2)} & a_{12}^{(2)} & \cdots & a_{1n}^{(2)} \\ 0 & a_{22}^{(2)} & \cdots & a_{2n}^{(2)} \\ \vdots & \vdots & & \vdots \\ 0 & a_{n2}^{(2)} & \cdots & a_{nn}^{(2)} \end{pmatrix} \tag{1.2.12}$$

avec $a_{11}^{(2)} = \alpha$ et $|\alpha| = \|a\|_2 \neq 0$ car la matrice $A^{(1)}$ est régulière par hypothèse. Dans cette première étape tous les éléments de la première colonne de $A^{(1)}$ sont annulés à l'exception de son premier élément qui concentre tout le "poids" (la norme euclidienne) du vecteur associé à la première colonne. Enfin les colonnes de $A^{(2)}$, à partir de la deuxième, sont calculées par les formules (1.2.10).

Soit $A_{22}^{(2)}$ la sous-matrice $(a_{ij}^{(2)}), 2 \leq i, j \leq n$, de $A^{(2)}$. Comme $\det(H^{(1)}) = -1$ on a donc $\det(A^{(2)}) = -\det(A^{(1)}) \neq 0$ car la matrice $A^{(1)}$ est régulière. D'autre part $\det(A^{(2)}) = a_{11}^{(2)} \det(A_{22}^{(2)})$ et les assertions précédentes impliquent que $\det(A_{22}^{(2)}) \neq 0$. La matrice $A_{22}^{(2)}$ est donc régulière et le processus de triangulation peut se poursuivre sur cette matrice en considérant maintenant sa première colonne. On voit que les éléments de la première ligne de $A^{(2)}$ ne sont plus concernés par la suite de la triangulation. L'examen du passage de $A^{(r-1)}$ à $A^{(r)}$ est fait dans le livre avec toutes les formules requises.

L'analyse matricielle de la méthode de Householder est particulièrement simple. On a

$$A^{(n)} = H^{(n-1)} \times H^{(n-2)} \times \cdots \times H^{(1)} \times A^{(1)}, \tag{1.2.13}$$

avec $A^{(n)}$ triangulaire supérieure. La matrice $A^{(n)}$ est inversible si $A = A^{(1)}$ est inversible.

Remarque 1.2.10. *Si $A = A^{(1)}$ n'est pas inversible, on trouvera un vecteur nul à une étape r ; ce vecteur étant le premier vecteur colonne d'une des matrices $A_{22}^{(r)}$. Dans ce cas, le r^{ieme} terme de la diagonale est un zéro, on prend $H^{(r)} = I$ et on peut continuer la triangulation jusqu'à la fin. La matrice $A^{(n)}$ a alors la configuration de la Figure 1.7.*

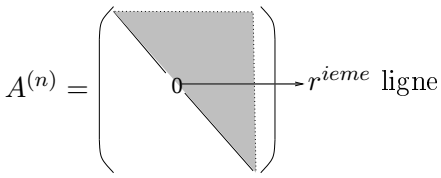


Figure 1.7:

Remarque 1.2.11. *La méthode de Householder est "stable" numériquement, au sens qu'elle ne dégrade pas le conditionnement initial de la matrice A. Elle utilise en effet des matrices de Householder H orthogonales et on sait (Théorème 1.2 des compléments en ligne) que le conditionnement* $\mathrm{cond}_2(A)$ *est invariant par transformation unitaire.*

Remarque 1.2.12. *Comme pour la méthode de Gauss (voir la Remarque 1.2.4), la méthode de Householder est un outil algorithmique d'évaluation du déterminant d'une matrice. Il est clair, d'après (1.2.13), que l'on a :*

$$[H^{(1)}]^{-1} \times \cdots \times [H^{(i)}]^{-1} \times \cdots \times [H^{(n-1)}]^{-1} A^{(n)} = A^{(1)} = A, \qquad (1.2.14)$$

comme l'inverse d'une matrice orthogonale de déterminant -1 *est une matrice orthogonale de déterminant* -1*, il vient :*

$$\det(A) = (-1)^{n-1} \det(A^{(n)}),$$

dès lors

$$\det(A) = (-1)^{n-1} \left(a_{11}^{(2)} \times \cdots \times a_{n,n}^{(n+1)} \right). \qquad (1.2.15)$$

L'algorithme de Householder donne une démonstration de la décomposition, dite "QU", d'une matrice.

Théorème 1.2.2. *Pour toute matrice A, il existe une matrice orthogonale Q, produit de* $(n-1)$ *matrices orthogonales élémentaires (les matrices de Householder), et une matrice triangulaire supérieure U, telles que* $A = QU$*.*

Preuve : Elle est facile dans le contexte. On a d'après (1.2.14)

$$A = A^{(1)} = [H^{(1)}]^{-1} \times \cdots \times [H^{(n-1)}]^{-1} A^{(n)}.$$

Posons $Q = [H^{(1)}]^{-1} \times \cdots \times [H^{(i)}]^{-1} \times \cdots \times [H^{(n-1)}]^{-1}$, nous savons (voir les propriétés de H à la Section 1.2.3) que $[H^{(i)}]^{-1} = H^{(i)}$ (avec éventuellement $H^{(i)} = I$), alors $Q = H^{(1)} \times \cdots \times H^{(i)} \times \cdots \times H^{(n-1)}$; posons de plus $U = A^{(n)}$ qui est triangulaire supérieure, la conclusion suit. □

Remarque 1.2.13. *Comme pour la méthode de Gauss on peut compter exactement le nombre d'op. él. requises pour la mise en œuvre de la méthode de Householder. On vérifie qu'il faut de l'ordre de* $2n^3/3$ *additions et de*

$2n^3/3$ *multiplications pour son fonctionnement, elle est donc deux fois plus chère que la méthode de Gauss. Notons aussi qu'elle nécessite le calcul de* n *racines carrées. Cependant la méthode ne demande aucune stratégie de pivot, on observe que son principe évite même toute idée de pivotement. Or rappelons que la recherche du pivot est nécessaire dans la méthode de Gauss, ce qui a évidemment un coût non négligeable. Affirmer que Householder est deux fois plus chère que Gauss est donc à nuancer.*

Intérêts de cette méthode Pour complèter la dernière remarque il est utile de récapituler ses avantages.

- C'est une méthode "stable" au sens où nous l'avons entendu à la Remarque 1.2.11.
- Par cette méthode on peut toujours trianguliariser une matrice quelconque, singulière ou non.
- Si on construit une matrice H telle que, tout vecteur $a \in \mathbb{R}^n$ peut être transformé en une combinaison linéaire de la forme $Ha = \alpha e_1 + \beta e_2$, où e_1 et e_2 sont les deux premiers vecteurs de la base canonique de $\mathbb{R}^n$, alors on peut transformer A, en la prémultipliant par H, en une matrice du type Hessenberg supérieur. L'algorithme de construction d'une telle matrice H utilise largement les idées précédentes et il est à peine plus compliqué.
- Par cette méthode on peut dès lors tridiagonaliser une matrice. Il suffit de pré- et de post-multiplier les matrices $A^{(r)}$ successives par des matrices convenables (qui transforment tout vecteur a en la somme $\alpha e_1 + \beta e_2$), on imagine assez naturellement le procédé !

Mentionnons que les deux dernières transformations de la matrice A sont utilisées dans certains algorithmes de calcul des valeurs propres de A. C'est une étape préalable à ce calcul.