Introduction

Pour ce devoir, il était demandé d'écrire des fonctions en langage Python permettant de résoudre éfficacement un système d'équations linéaires issu d'un modèle d'éléments finis appelé `ccore`. Le devoir se divise en deux tâches : d'abord la résolution d'un système en format plein et puis en format creux.

1 Format plein

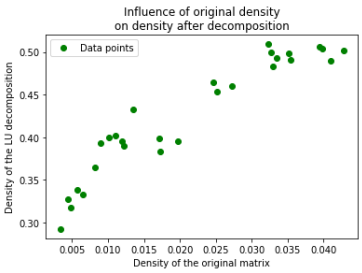
Dans la partie en format plein du devoir, le but était d'étudier la densité de la factorisation LU en place d'une matrice $A \in \mathbb{C}^{n\times n}$ de faible densité. Pour faire cela, il a fallu écrire une fonction capable de faire cette factorisation : `LUfactorize`. Elle prend en argument la matrice A et décompose celle-ci en un produit d'une matrice triangulaire supérieure U et d'une matrice triangulaire inférieure à diagonale unité L de sorte à ce que $A = LU$.

Pour résoudre le système, il suffit alors de d'abord calculer $Ly = b$ par substitution avant pour trouver y, puis de résoudre $Ux = y$ par substitution arrière pour trouver x. `LUsolve` résout ainsi le système $Ax = b$.

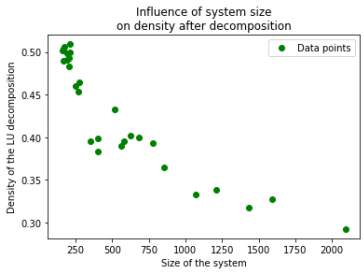
1.1 Étude de la densité de la décomposition LU

Comme la matrice A est issue d'un modèle d'éléments finis, elle contient beaucoup d'entrées nulles. La densité d'une matrice est définie comme $\operatorname{dens} A = \frac{\operatorname{nnz}}{n^2}$, où nnz est défini comme le nombre d'éléments non nuls de A. On note la densité de A après factorisation en place comme $\operatorname{dens}(LU)$ par simplicité.

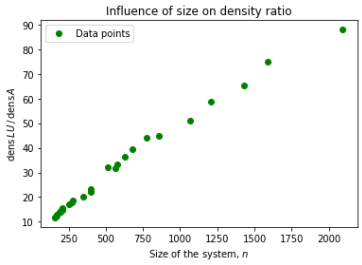
Afin de faire une étude correcte de l'évolution de la densité de la matrice A après sa décomposition LU en place, plusieurs graphes ont été réalisés. Ils sont représentés à la Figure 1.



(a) Valeur de $\operatorname{dens}(LU)$ en fonction de $\operatorname{dens} A$.



(b) Influence de la taille du système sur $\operatorname{dens}(LU)$.



(c) Rapport entre $\operatorname{dens}(LU)$ et $\operatorname{dens} A$ en fonction de n.

FIGURE 1 – Différents graphes pertinents pour l'analyse de densité en section 1.1.

On remarque que plus la matrice devient grande, plus elle devient creuse ; c'est une propriété du modèle d'éléments finis. Sur le graphe de la Figure 1a, on remarque que la densité de la factorisation augmente lorsque la densité de A augmente. Cependant, comme expliqué à la section 2.3, ce rapport n'est que qualitatif et dépend du contenu de la matrice A.

En combinant le fait que la densité de A diminue plus le système devient grand, et que la densité de la factorisation diminue lorsque la densité de A devient plus petite, on peut prédire par transitivité que plus le système devient grand, plus la densité de sa factorisation LU va diminuer. C'est ce qu'on observe sur le graphe de la Figure 1b. Finalement, le rapport de la densité de la matrice après factorisation sur sa densité avant factorisation est représenté comme une fonction de la taille du système à la Figure 1c.

1.2 Complexité temporelle de `LUsolve`

Le solveur `LUsolve` fait deux choses séquentiellement, et la complexité totale est donc la somme des deux complexités partielles.

1.2.1 Complexité temporelle de `LUfactorize`

Il s'agit ici de calculer le nombre d'opérations pour une décomposition LU en place. Pour cela, on se réfère au livre de référence [1] pp. 151–152. Le nombre d'opérations de l'Algorithme 20.1 est dominé par l'opération vectorielle $u_{j,k:n} = u_{j,k:n} - \ell_{jk} u_{k,k:n}$. Soit $l = n-k+1$ la longueur des vecteurs lignes étant manipulés. On voit que l'opération dominante effectue une multiplication scalaire-vecteur et une soustraction entre vecteurs. Le

1. Trefethen, Lloyd N. & Bau, David III. (1997). *Numerical Linear Algebra*. Philadelphia : PA, SIAM.

nombre de *flops* est alors de $2l$. Pour chaque valeur de k, l'indice de la boucle extérieure, la boucle intérieure est répétée pour toutes les lignes de $k+1$ jusque n. La complexité de `LUfactorize`, $d(n)$, est donc

$$d(n) \sim \sum_{k=1}^{n} \underbrace{(n-k)}_{j} 2(n-k+1) = 2\sum_{j=0}^{n-1} j(j+1) = 2\sum_{j=0}^{n-1} j^2 + 2\sum_{j=0}^{n-1} j = 2\left(\frac{n^3}{3} - \frac{n^2}{2} + \frac{n}{6}\right) + 2\left(\frac{n^2}{2} - \frac{n}{2}\right) \sim \frac{2n^3}{3}\,.$$

1.2.2 Complexité temporelle des substitutions

On sait que la substitution arrière effectue $\frac{n^2-n}{2}$ soustractions, $\frac{n^2-n}{2}$ multiplications et n divisions. Pour la substitution avant, les opérations effectuées sont $\frac{n^2-n}{2}$ soustractions et $\frac{n^2-n}{2}$ multiplications. Pour obtenir ces résultats, calculons la complexité temporelle des substitutions, $\aleph(n)$, qui est alors la somme des complexités de chacune, c'est-à-dire

$$\aleph(n) \sim \underbrace{\sum_{k=1}^{n} 2(\overbrace{n-k}^{j})}_{\text{avant}} + \underbrace{\sum_{k=1}^{n} \big(2(n-k)+1\big)}_{\text{arrière}} \sim 2\sum_{j=0}^{n-1} j + 2\sum_{j=0}^{n-1} j + n = 4\left(\frac{n^2}{2} - \frac{n}{2}\right) + n \sim 2n^2\,.$$

1.2.3 Complexité totale

Comme dit plus haut, pour trouver $s(n)$, la complexité temporelle totale, il suffit de sommer les deux complexités (comme les opérations se font en série). On trouve alors

$$s(n) = d(n) + \aleph(n) \sim \frac{2n^3}{3} + 2n^2 \sim \frac{2n^3}{3}\,.$$

Comme on peut le voir, l'étape de factorisation est dominante pour la complexité temporelle. On voit sur le graphe de la Figure 2 que cette prédiction théorique est confirmée expérimentalement. Mentionnons finalement que comme les matrices sont issues d'un modèle d'éléments finis, elles sont symétriques (hermitiennes) et définies positives ; il n'est alors pas nécessaire de rajouter une vérification pour éviter les divisions par zéro dans `LUfactorize`. Ceci se démontre par le critère de Sylvester.

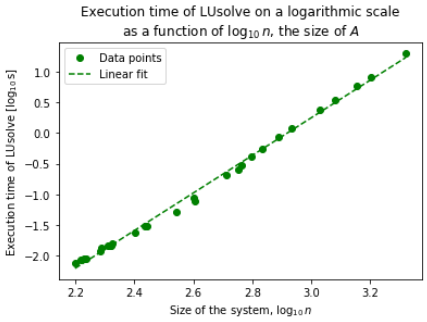


FIGURE 2 – Temps d'exécution de `LUsolve` en fonction de n. La pente de 3 montre la relation cubique.

2 Format creux

Les matrices issues de `ccore` sont très creuses. Il serait donc intéressant de trouver une manière de résoudre le système sans avoir à faire autant de calculs. Pour faire cela, nous utilisons le format CSR et la fonction `CSRformat` afin avoir une représentation de la matrice qui n'utilise pas autant de mémoire. Une fois celle-ci obtenue, il est possible d'améliorer la complexité de la résolution du système par factorisation LU.

2.1 Performances de `CSRformat`

La fonction `CSRformat` parcourt toute la matrice, ligne par ligne, en copiant toutes les entrées non nulles dans un vecteur `sA`, leurs indices de colonne dans `jA` et l'indice de leur premier élément non nul dans `iA`. On voit donc facilement que cette fonction doit avoir une complexité temporelle de $\Theta(n^2)$. Les opérations les plus coûteuses effectuées sont le parcours de la matrice pour trouver la valeur de nnz en $\Theta(n^2)$ et la recherches d'indices et de valeurs d'éléments non nuls sur une ligne, en $\Theta(n^2)$ au total pour n lignes. La complexité totale est donc également en $\Theta(n^2)$, ce qui se voit sur le graphe de la Figure 3.

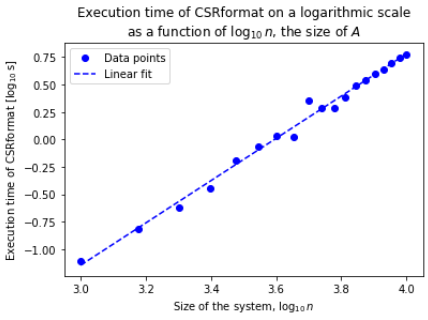


FIGURE 3 – Temps d'exécution de `CSRformat` en fonction de n. La pente de 2 montre la relation quadratique.

2.2 Complexité temporelle de `LUcsrsolve`

La fonction `LUcsrsolve` est une version optimisée de `LUsolve` spécifiquement pour les matrices creuses. Il est important pour les solveurs creux de connaître la bande de la matrice A. Définissons la largeur de bande k de A comme étant

$$k = \max(k_1, k_2)\,, \quad \text{où} \quad a_{ij} = 0 \quad \text{si} \quad j < i - k_1 \quad \text{ou} \quad j > i + k_2\,; \quad k_1, k_2 \geq 0\,.$$

En effet, une des propriétés de la factorisation LU est que lors de la décomposition, des nouvelles entrées non nulles peuvent apparaître mais uniquement à l'intérieur de la bande de la matrice (phénomène de *fill-in*). On peut donc préallouer facilement cette bande, la remplir au fur et à mesure de la factorisation, et n'avoir qu'à considérer une partie de cette bande (et donc de la matrice A) lors des substitutions. Pour calculer la complexité de la fonction `LUcsrsolve`, il faut observer qu'elle sera égale à la somme des complexités de `LUcsr` et de deux substitutions modifiées.

2.2.1 Complexité temporelle de `LUcsr`

La fonction `LUcsr` a une complexité qui dépend de plusieurs étapes : d'abord le calcul de la largeur de bande en $\Theta(n)$ grâce au format CSR en utilisant le vecteur `iA` ; ensuite, la préallocation de vecteurs de taille au plus nk en $\mathcal{O}(nk)$ et finalement la réelle décomposition LU creuse en $\lambda(n,k)$. Pour trouver $\lambda(n,k)$, on observe que la boucle intérieure fait deux opérations : (une multiplication et une soustraction), et on calcule (en idéalisant un peu)

$$\lambda(n,k) \in \mathcal{O}\left(\sum_{i=0}^{n-1} \sum_{j=i+1}^{\min(i+k_1+1,n)-1} \sum_{m=i+1}^{\min(i+k_2+1,n)-1} 2\right) \subseteq \mathcal{O}\left(2\sum_{i=0}^{n-1} k_1 k_2\right) \subseteq \mathcal{O}(2nk_1k_2) \subseteq \mathcal{O}(nk^2)\,.$$

2.2.2 Complexité des substitutions modifiées

Dans les substitutions modifiées, on peut utiliser le fait que seule la partie non nulle de la matrice nous intéresse. On écrit alors la complexité $\sigma(n,k)$ des deux substitutions en série comme

$$\sigma(n,k) \sim \underbrace{\sum_{i=0}^{n} \sum_{j=\max(0,i-k_1+1)}^{i-1} 1}_{\text{avant}} + \underbrace{\sum_{i=0}^{n} \sum_{j=i+1}^{\min(i+k,n)-1} 1}_{\text{arrière}} \sim nk_1 + nk_2 \sim 2nk \in \mathcal{O}(nk)\,.$$

2.2.3 Complexité totale de `LUcsrsolve`

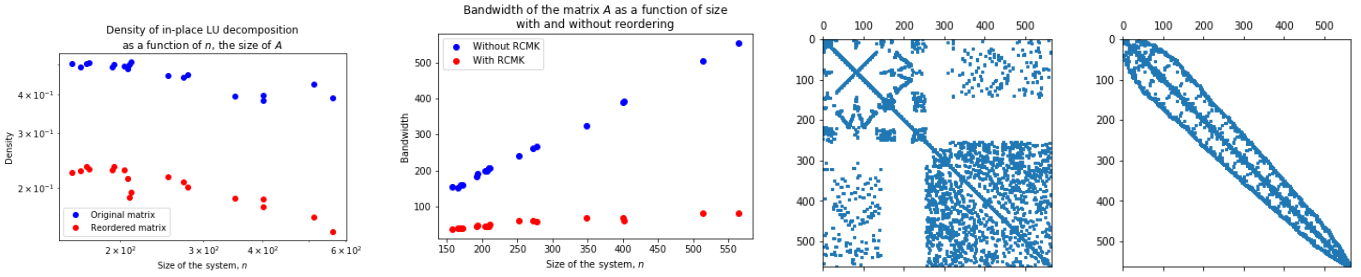
Comme dit plus haut, ces opérations se font toutes en série, et la complexité totale $t(n,k)$ s'écrit alors

$$t(n,k) \in \mathcal{O}(nk^2 + nk) \subseteq \mathcal{O}(nk^2)\,.$$

La factorisation est donc l'étape contribuant le plus à la complexité temporelle.

2.3 Numérotation « reverse Cuthill-McKee »

Comme expliqué à la section 2.2, la complexité de la fonction `LUcsrsolve` dépend du carré de la largeur de bande. Il serait donc intéressant d'utiliser un algorithme de renumérotation des nœuds tel que RCMK pour diminuer cette largeur de bande. Cet algorithme n'influence pas uniquement la complexité en temps mais aussi en mémoire : lors du *fill-in* de la décomposition LU, plus la bande est petite, plus on peut borner la densité de la matrice résultant de la factorisation en place.



(a) Comparaison entre la densité de la factorisation avec et sans RCMK.

(b) Influence de RCMK sur la largeur de bande de A.

(c) Matrice A sans renumérotation RCMK.

(d) Matrice A avec renumérotation RCMK.

FIGURE 4 – Influence de la renumérotation RCMK sur la densité de A.

Sur la Figure 4, on peut voir que la densité de la factorisation LU diminue lorsque la matrice est permutée (Figure 4a) et que la largeur de bande de la matrice A diminue (et augmente beaucoup moins rapidement) lorsque la matrice est permutée (Figure 4b). Les Figures 4c et 4d donnent une vue qualitative.

2.4 Complexité de `LUcsrsolve` avec renumérotation RCMK

Afin d'évaluer l'utilité de cette renumérotation, il est intéressant de regarder l'évolution de la complexité lorsqu'on applique la permutation. Il faut alors rajouter deux étapes à la fonction : le calcul de la permutation optimale (fonction `RCMK`) et son application.

2.4.1 Complexité du calcul de la permutation

Soit $M \in \{0,1\}^{n\times n}$ la matrice telle que $m_{ij} = 1$ si $a_{ij} \neq 0$ et 0 sinon, et soit $G(E,V)$ le graphe dont M est la matrice d'adjacence, avec E l'ensemble des arêtes et V l'ensemble des nœuds. L'étape dominante pour la complexité temporelle est le fait de devoir trier à chaque itération la liste des nœds adjacents au nœud venant de sortir de la file. Or nous pouvons borner la taille de cette liste de nœuds adjacents par le degré maximal dans le graphe G, $\Delta(G)$. On remarque alors que pour toute permutation de la matrice (toute renumérotation des nœuds), et donc en particulier pour la permutation RCMK, le degré maximal est inférieur au double de la largeur de bande augmenté de 1 (pour compter l'élément diagonal) : $\Delta(G) \leq k_1 + k_2 + 1 \leq 2k + 1$. En supposant qu'un algorithme de tri avec une complexité dans le pire cas linéarithmique, tel que *mergesort*, soit utilisé, on peut alors dire que la fonction `RCMK` a une complexité $r(G)$ en $\mathcal{O}\big(|V|\Delta \log \Delta\big)$, ou bien, avec les paramètres de notre système RCMK, n et k,

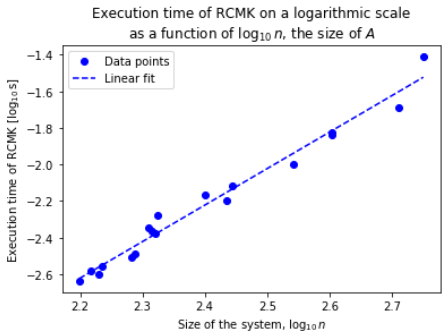


FIGURE 5 – Temps d'exécution pour le calcul de la permutation RCMK.

$$r(n,k) \in \mathcal{O}\big(n(2k+1)\log(2k+1)\big) \subseteq \mathcal{O}(nk\log k)\,.$$

Le temps d'exécution de la fonction `RCMK` est montré sur la Figure 5.

2.4.2 Complexité de l'application de la permutation

Pour permuter la matrice, on peut imaginer qu'il faut permuter d'abord les lignes puis les colonnes. Pour les lignes, il faut donc déplacer au plus $2nk$ éléments (comme chacune des n lignes a au plus de l'ordre de $2k$ entrées), et pour les colonnes la même chose. Il est alors possible d'appliquer la permutation en $\mathcal{O}(nk)$.

2.4.3 Complexité de `LUcsrsolve` avec renumérotation RCMK

La complexité totale $\tau(n,k)$ pour `LUcsrsolve` avec renumération est alors (en comptant toujours la complexité de `LUcsr`)

$$\tau(n,k) \in \mathcal{O}(nk^2 + nk\log k + nk) \subseteq \mathcal{O}(nk^2)\,, \quad \text{car} \quad \lim_{(n,k)\to(+\infty,+\infty)} (nk^2 + nk\log k) = \lim_{(n,k)\to(+\infty,+\infty)} nk^2\,.$$

C'est la même complexité que pour `LUcsrsolve` sans renumérotation, mais on remarque que comme la largeur de bande a fortement diminué, la résolution du système sera beaucoup plus rapide pour les matrices creuses. C'est aussi ce qu'on observe sur la Figure 6. On remarque que le solveur plein est plus rapide pour les matrices de ces tailles (grâce à la vectorisation et aux routines optimisées de `BLAS`), mais la pente pour le solveur avec RCMK est plus petite, et à partir d'un certain $n_0 \gg 0$, la résolution avec permutation sera plus rapide.

Un autre point intéressant est que dans le cas de matrices très mal numérotées ($k \approx n$), l'algorithme creux devient un simple algorithme de factorisation en $\mathcal{O}(n^3)$ et sa pente est égale à la pente pour le solveur plein.

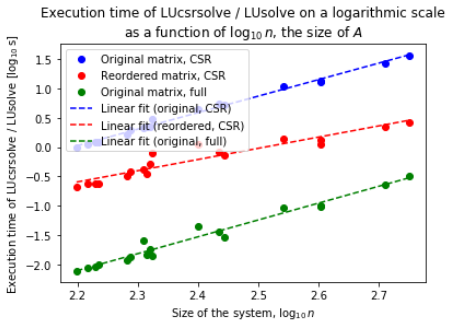


FIGURE 6 – Temps d'exécution pour les différents solveurs.

Conclusion

La résolution de systèmes algébriques est un aspect important de l'analyse numérique, et la décomposition LU est souvent le meilleur choix pour le faire rapidement. Malgré la simplicité de l'algorithme, il permet de résoudre rapidement des systèmes de grande taille. Pour les matrices issues d'un modèle d'éléments finis, on sait qu'elles sont creuses et hermitiennes. Il est alors possible de les réordonner par une algorithme tel que RCMK afin de diminuer le temps pour résoudre le système en profitant de l'aspect « bande » de telles matrices.