PROJET D'ALGORITHMIQUE

Complexes simpliciaux pour l'analyse topologique des données

Décembre 2024 - Février 2024

Ousema Bouaneni, Pierre Sara

École Polytechnique | Institut Polytechnique de Paris

TABLE DES MATIÈRES

1
LES COMPLEXES DE CECH ET LE PROBLÈME DE BOULE MINIMALE ENGLOBANTE

1.1 - Le problème de boule minimale englobante

Le problème originel de recherche de la boule minimale englobante est assez simple : il s'agit de trouver une boule de rayon minimal, si elle est existe, englobant les points d'un ensemble P donné.

On remarque assez rapidement par un argument de compacité sur la fonction $x \to \sup_{x\in P}\|x-p\|$ que cette boule existe et est unique dès que P est fini. On notera donc cette boule $\mathbf{meb}(\boldsymbol{P})$.Toutes les démonstrations nécessaires sont en annexe.

1.1.1 • Algorithme et correction

1.1.1.1 - L'algorithme

Une première idée d'algorithme est due à Welzl :

WELZL_AUX (P,R):
if $P = \emptyset$ **or** $\#R = d+1$:
    D := trivial($\emptyset$,R)
else :
    choisir uniformément $p \in P$
    $D = \textsc{welzl\_aux}(P \setminus \{p\}, R)$
    if $p \in D$:
        return D
    $D = \textsc{welzl\_aux}(P \setminus \{p\}, R \cup \{p\})$
return D

WELZL_MEB (P):
1 **return** WELZL_AUX$(P, \emptyset)$

La terminaison est ici trivialement assurée par la décroissance de la fonction $f : P, R \to d + 1 - \#R$ à valeurs entières positives. La preuve de la correction, classique, est en annexe.

1.1.2 • Le problème du cas de base

Reste à trouver une fonction efficace pour effectuer le cas de base.

1.1.2.1 - Transformation du problème en problème matriciel

On s'appuie ici sur le fait démontré en annexe que le centre de la sphère minimale, s'il existe, est nécessairement dans le plan affine engendré par les points de P.

Fixons ici $R = \{p_0, ..., p_n\}$ avec $n \leq d + 1$. La recherche de $\mathrm{MB}(\emptyset, R)$ repose sur le résultat suivant :

La correction du cas récursif et l'existence de $\mathrm{meb}(P)$ montre qu'on peut fixer c le centre et r le rayon de $\mathrm{MB}(\emptyset, R)$. On peut alors exprimer c en coordonnées barycentriques : il existe $\lambda_0, ..., \lambda_n \in \mathbb{R}$ tels que $\sum_{i=0}^{n} \lambda_i = 1$ et $c = \sum_{i=0}^{n} \lambda_i p_i$.

En utilisant, pour $0 \leq j \leq n$, la relation $\|\overrightarrow{cp_j}\|^2 = r^2$ et en la développant, il vient $2 \sum_{i=0}^{n} \left(\overrightarrow{Op_i} \mid \overrightarrow{Op_i}\right) + \alpha = - \|\overrightarrow{Op_i}\|^2$ avec $\alpha = \|\overrightarrow{Oc}\|^2 - r^2$.

Avec $\sum_{i=0}^{n} \lambda_i = 1$, on trouve le problème matriciel $AX = B$ avec :

- $A = \begin{pmatrix} 2P^TP & 1 \\ 1 & 0 \end{pmatrix}$ et $P = \mathrm{Mat}\left(\overrightarrow{Op_0}, ..., \overrightarrow{Op_n}\right)$
- $X = \begin{pmatrix} \lambda_0 \\ \vdots \\ \lambda_n \\ \alpha \end{pmatrix}$
- $B = \begin{pmatrix} -\|\overrightarrow{Op_0}\|^2 \\ \vdots \\ -\|\overrightarrow{Op_0}\|^2 \\ 1 \end{pmatrix}$

Une analogie avec la matrice de Gram permet de montrer (cf. annexe) que A est inversible si les points sont affinement indépendants.

Cette condition est vérifiée dans le cas de l'algorithme présent, comme montré en annexe.

1.1.3 • Une autre vision du problème et algorithme de Seidel

Cette vision géométrique se généralise en tant que problème de type LP, ce qui fait qu'on peut utiliser d'autres algorithmes que celui de Welzl et ce dans un cadre plus général.
La comparaison entre l'algorithme développé précédemment et l'algorithme de Welzl général nous fait facilement intuiter que :

- l'ensemble des contraintes S est l'ensemble des points P de départ
- la fonction f d'évaluation associe pour tout $P' \subset P$ le rayon de $\mathrm{meb}(P')$.
- Les bases correspondent aux ensembles $P' \subset P$ tels que $\mathrm{meb}(P') = \mathrm{MB}(\emptyset, P')$ et on a vu par lemme que son cardinal peut être majoré par $d + 1$ qui est donc la dimension combinatoire du problème

Les propriétés de monotonie et de localité découlent directement des lemmes montrés précédemment.
On peut alors résoudre ce problème par un solveur général de problèmes de type LP utilisant l'algorithme de Welzl ou celui de Seidel.

1.2 - Application aux complexes de Cech

1.2.1 • Version naïve

La version naïve s'appuie sur le résultat suivant :

Lemme

Soit $P = \{p_0, ..., p_n\}$ un ensemble de points. Alors le simplexe C formé à partir des points de P est un complexe de Cech de valeur de filtration le rayon de $\mathrm{meb}(P)$.

Démonstration

Soit $r > 0$. P est un complexe de Cech de valeur de filtration $\leq r$ ssi il existe $c \in \cap_{p \in P}$ $B(p, r)$ ssi $B(c, r)$ est une boule englobante de P ssi $r \geq r^*$ où r^* est le rayon de $\mathrm{meb}(P)$.

Pour afficher tous les simplexes avec leurs valeurs de filtration, on énumère tous les sous-ensembles de P et on construit leurs boules minimales englobantes.

1.2.2 • Version améliorée

Si on cherche tous les simplexes de valeur de filtration inférieure à une certaine limite l, on peut alors remarquer que celle-ci est croissante pour l'inclusion par unicité de la boule minimale englobante. On peut donc modifier l'énumération des sous-ensembles de P pour procéder par cardinal croissant : à chaque étape, on garde en mémoire les simplexes de taille k dont les valeurs de filtration sont restées inférieures à la limite, et on passe aux simplexes de taille $k + 1$ en ajoutant un élément et en recalculant les valeurs de filtration.

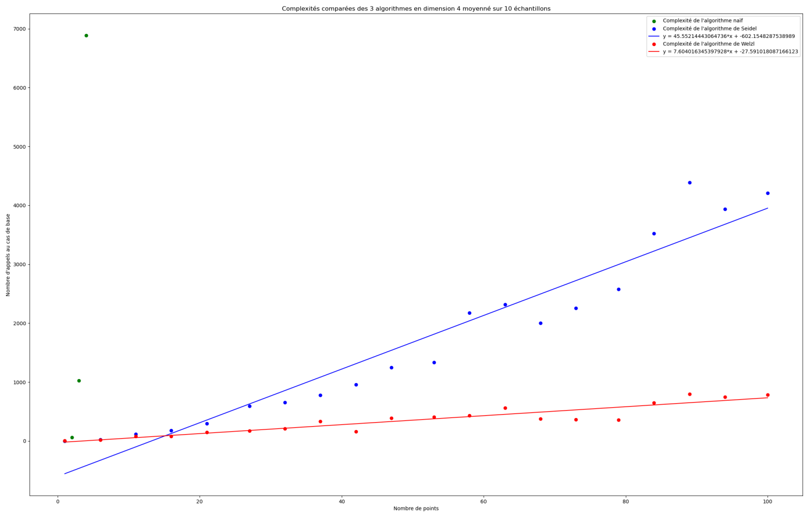
1.3 - Stratégie de test

On a souhaité effectuer 3 types de tests :

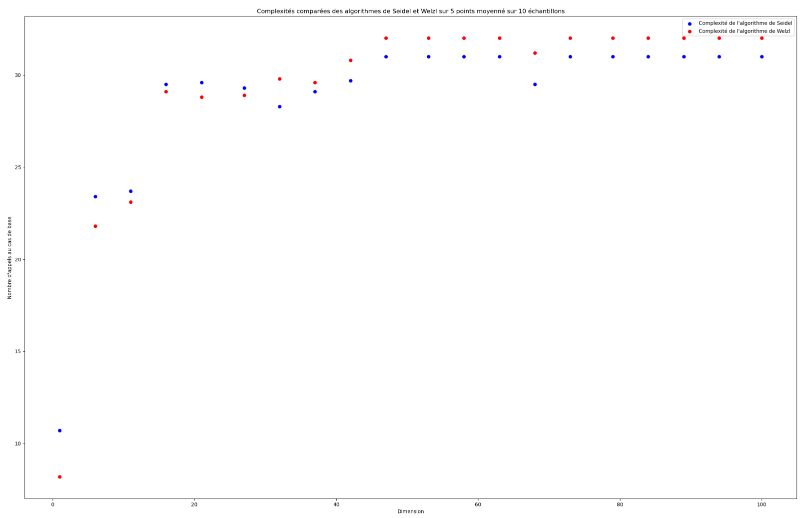
- On a tout d'abord fait des tests manuels en faible dimension et avec peu de points sur lesquels on connaît le résultat.
- On a ensuite effectué des tests randomisés de résultat prévisible : singletons, paires de points et points tirés uniformément dans une sphère, dans une boule ou dans un cube. Ceci permet de tester nos algorithmes en grande dimension.
- Enfin, on a effectué des tests aléatoires en comparant le résultat donné par notre algorithme avec celui d'un algorithme naïf en faible dimension (qui testait en l'occurrence tous les sous-ensembles de P de taille au plus $d + 1$).

1.3.1 • Complexité comparée

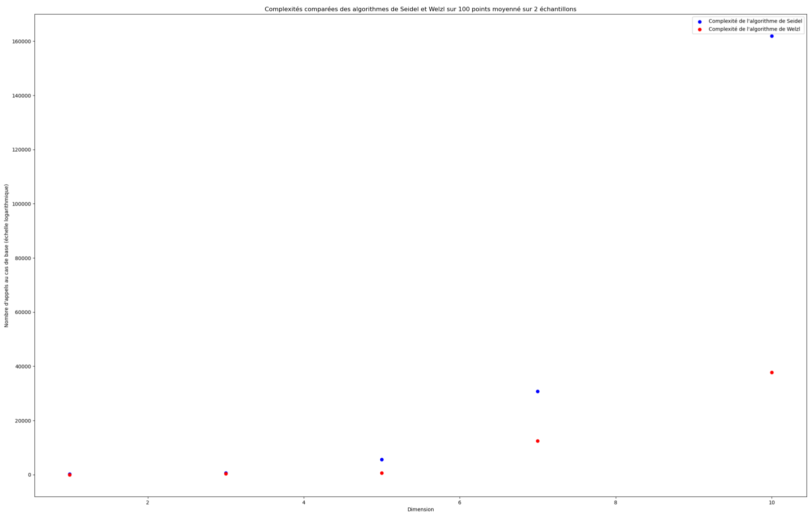
Quand on fixe la dimension et qu'on regarde la complexité en fonction du nombre de points seulement, le constat est assez clair : l'algorithme naïf est exponentiel et inefficace, tandis que les deux autres algorithmes ont alors une complexité visiblement linéaires avec une constante légèrement plus grande pour l'algorithme de Seidel que pour celui de Welzl:



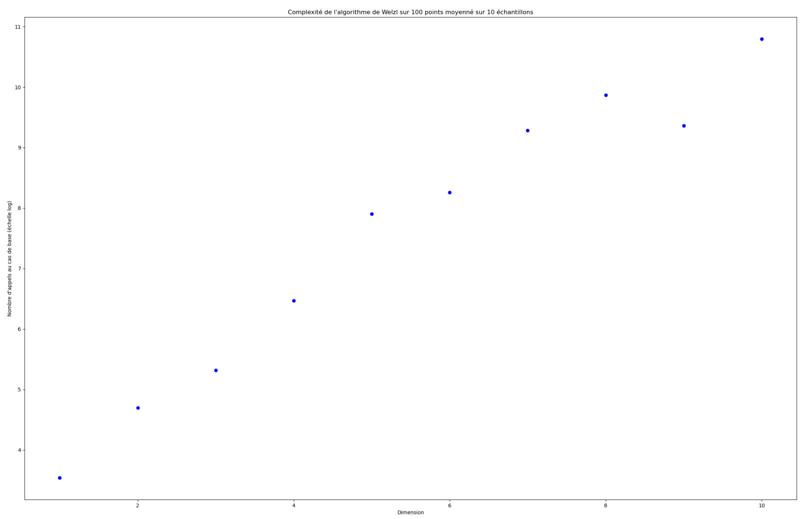
Quand on regarde la dépendance par-rapport à la dimension, c'est plus complexe : en effet, si le nombre de points fixe qu'on choisit est très petit devant la dimension, le cas de base bloquant est celui où $P = \emptyset$. On n'atteint alors pas la complexité factorielle mais on a une stabilisation à 2^d opérations pour Welzl et à $2^d - 1$ pour Seidel (ce qui correspond au pire des cas : on peut supposer qu'il est alors très probable que ce pire cas où chaque point est en-dehors de la sphère formée est fréquent).



En revanche, quand on prend une taille nettement plus grande, on retrouve alors la complexité exponentielle.



Le passage à l'échelle logarithmique montre qu'on est dans un cas visiblement légèrement sur-exponentiel :



2
LES α –COMPLEXES

2.1 - VERS UN NOUVEAU PROBLÈME DE BOULE MINIMALE ENGLOBANTE

2.1.1 • LES α –COMPLEXES

On définit ici $\mathrm{MB2}(P, R)$ comme une boule de rayon minimal possédant les points de R sur sa frontière et aucun point de P dans son intérieur si elle existe.
La preuve alternative utilisée pour l'unicité de $\mathrm{meb}(P)$ montre, en inversant les deux dernières inégalités par hypothèse sur P, l'unicité d'une telle boule si elle existe.

2.2 - RÉSOLUTION DU NOUVEAU PROBLÈME DE BOULE MINIMALE ENGLOBANTE

2.2.1 • ADAPTATION DE L'ALGORITHME DE WELZL

On peut adapter l'algorithme de Welzl pour ce nouveau problème. Le cas récursif s'appuie sur la même idée : on prend $p \in P$, on calcule récursivement $\mathrm{MB2}(P \setminus \{p\}, R)$. Si p n'est pas dans l'intérieur cette fois-ci, alors on a trivialement que $\mathrm{MB2}(P, R) = \mathrm{MB2}(P \setminus \{p\}, R)$. Sinon, on peut montrer que p est nécessairement sur la frontière.

MEB2 (P,R):
if $P = \emptyset$ **or** $\#R = d + 1$:
  $D :=$ trivial($\emptyset$,R)
else :
  choisir uniformément $p \in P$
  $D = \text{MEB2}(P \setminus \{p\}, R)$
  if $p \notin \text{Int}(D)$:
    return D
  $D = \text{MEB2}(P \setminus \{p\}, R \cup \{p\})$
return D

La démonstration de la correction (en annexe) se fait de façon très proche de celle de l'algorithme original.

Le cas de base est identique mais on est contraint d'extraire une famille de points affinement indépendante de taille maximale dans le cas où les points P ne le sont pas. Pour cela, on a adapté l'élimination de Gauss-Jordan en gardant trace des permutations : les colonnes non nulles donnent alors une famille maximale de vecteurs libres en les associant aux vecteurs originaux.

Cette version n'a malheureusement pas validé les tests.

2.2.2 • Utilisation d'un solveur de problèmes de type LP

Notre nouveau problème est aussi un problème de type LP. De fait, on peut intuiter encore une fois grâce à l'algorithme de Welzl que l'on peut prendre comme ensemble de contraintes $S = P \setminus R$, comme base initiale l'ensemble R et associer à chaque sous-ensemble A de S le rayon de la boule minimale contenant les éléments de $P \setminus A$ sur sa frontière et ceux de A dehors si une telle boule existe et $+\infty$ sinon. La monotonie est alors évidente et la localité découle des mêmes résultats que ceux qui montrent la correction de l'algorithme de Welzl adapté.

Ceci nous permet d'appliquer l'algorithme de Seidel à ce nouveau problème.