

# Chemin glouton orthogonal le plus long d'une grille

## Rapport de stage de L3

Hugo JACOB

juin-juillet 2020

### 1 Introduction

Ce stage a été réalisé en télétravail sous la direction de Cyril Gavoille du LaBRI, que je remercie de m'avoir proposé ce stage et de m'avoir guidé au cours des six semaines.

#### 1.1 Contexte général

Le sujet du stage s'inscrit dans l'étude du problème du voyageur de commerce, un célèbre problème d'optimisation. Etant donné un graphe pondéré, on cherche un cycle passant par tous les sommets de pondération minimale. Ce problème est connu pour sa difficulté. Le problème de décision associé est NP-complet et des hypothèses supplémentaires sont nécessaires pour approximer une solution. Si on suppose que les distances correspondent à une métrique euclidienne, le problème demeure NP-complet [Pap77]. Dans le cas métrique, on suppose que les distances vérifient l'inégalité triangulaire. Cette difficulté explique l'attention portée aux heuristiques et algorithmes d'approximation. Les plus connus sont l'heuristique de Lin-Kernighan qui obtient de bons résultats en pratique et l'algorithme de Christofides qui a un ratio d'approximation de  $\frac{3}{2}$  dans le cas métrique.

#### 1.2 Problème étudié

Le stage porte sur l'étude du pire cas de l'heuristique gloutonne du plus proche voisin. Elle consiste à se déplacer vers le sommet non parcouru le plus proche jusqu'à avoir parcouru tous les sommets. On s'intéresse plus précisément à une structure de graphe utilisée dans un article récent [HW15] qui montre une borne inférieure sur ce pire cas dans le cas métrique. De plus, on considérera

plutôt le problème du voyageur de commerce dans sa version avec un chemin car l'heuristique considérée s'y prête.

### 1.3 Résultats connus

Une borne supérieure a déjà été établie dans le cas métrique [RSLI77] : le ratio d'approximation pour un cycle est au plus de  $\frac{1}{2} (\lceil \log_2(n) \rceil + 1)$  où  $n$  est le nombre de sommets.

Le même article présente une borne inférieure dans le cas métrique : le ratio d'approximation pour un cycle est d'au moins  $\frac{1}{3} \log_2(n + 1) + \frac{4}{9}$ , avec une construction assez complexe. Une borne inférieure de  $\frac{1}{4} \log_2(n - 1)$  a été établie sur la structure de graphe étudiée avec une construction plus simple [HW15], après une première publication présentant une preuve fausse [BH15]. Elle améliore des résultats précédemment établis pour les métriques rectilinéaire et euclidienne (distances associées aux normes 1 et 2 respectivement).



Dans cette partie, on essaie d'améliorer la construction proposée dans [HW15]. Commençons par la présenter :

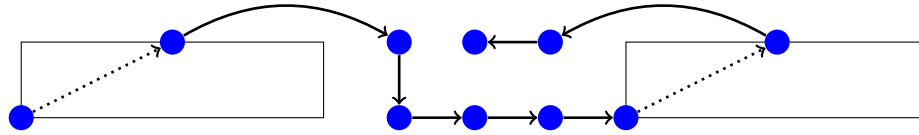


FIGURE 2 – Motif de construction

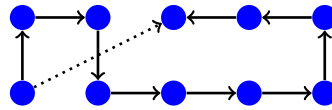


FIGURE 3 – Motif initial

Le motif de construction illustre comment produire un motif deux fois plus gros en dupliquant un premier motif qu'on place au niveau des rectangles. On voit apparaître les deux sauts qui couvrent la moitié de la longueur de la grille, ce qui explique le comportement asymptotique de la construction que nous verrons plus tard.

Quelques invariants doivent être respectés pour que la construction produise un chemin glouton valide. Il faut que le point initial soit sur une extrémité et que le point final soit centré, mais aussi que le chemin glouton produit soit toujours valide si on ajoute des points non visités sur les côtés de la grille comme illustré sur la figure qui suit.

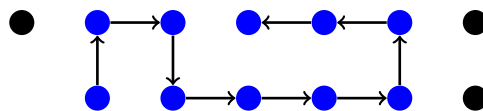


FIGURE 4 – Contraintes sur le motif initial

On peut améliorer légèrement la construction comme suit :

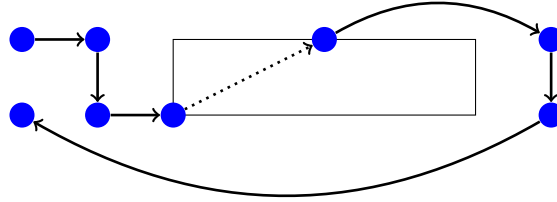


FIGURE 5 – Ajout final au motif

On ne peut alors plus agrandir le motif mais on gagne une longueur significative en ajoutant seulement 3 colonnes.

Il est également possible de changer de motif initial tant qu'on respecte les contraintes mentionnées. On peut remarquer que la longueur du motif doit impérativement être impaire. Pour trouver des motifs efficaces, on effectue des recherches exhaustives sur les chemins gloutons orthogonaux respectant les contraintes mentionnées précédemment.

**Proposition 2.1** *Avec un motif initial de longueur  $A$  sur une grille de dimension  $2 \times D$ . La construction d'ordre  $k$  améliorée produit un chemin glouton orthogonal d'une grille  $2 \times (2^k(D + 3))$ , de longueur  $2^{k-1} (2 * A + (k + 3)(D + 3) + 6) - 1$ .*

Les graphiques suivants permettent de comparer les différents motifs calculés. L'ordre vertical des points du second graphique correspond à l'ordre des asymptotes. Plus précisément, la valeur en ordonnée sur ce graphique est le coefficient du terme prépondérant (parmi les termes distincts) de l'expression de la longueur du chemin.

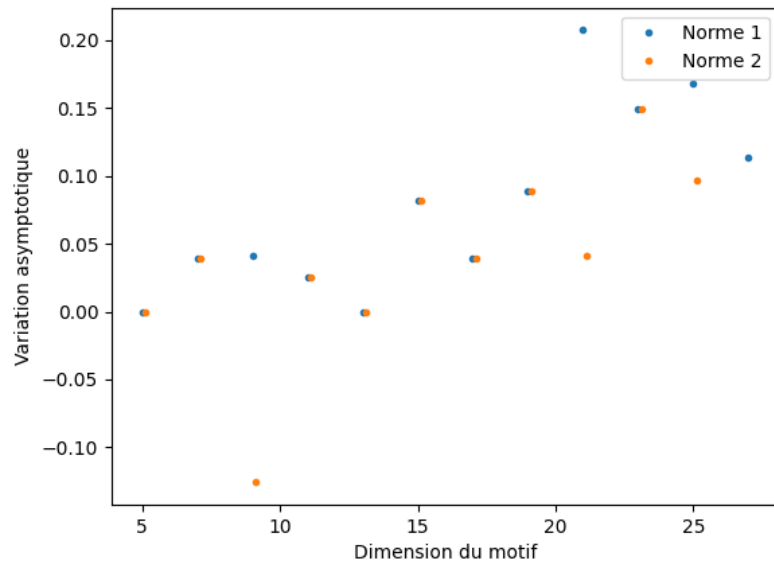
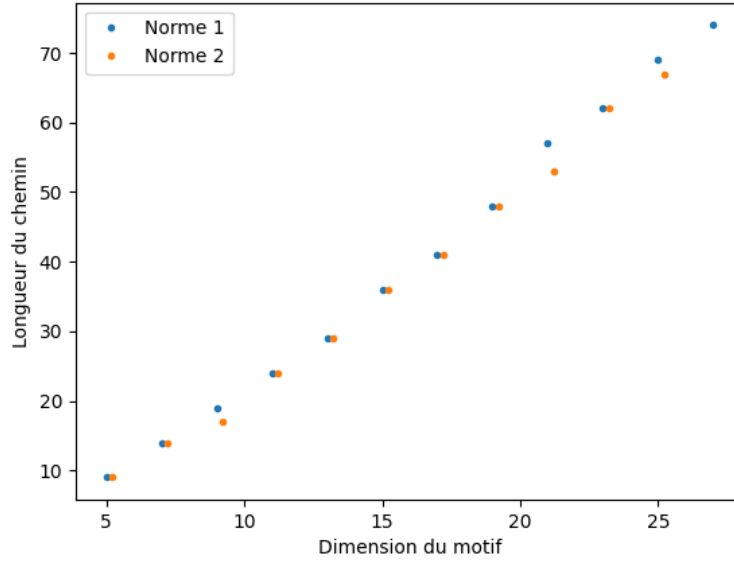


FIGURE 6 – Efficacité des motifs calculés

On en déduit que le motif initial de dimension  $2 \times 21$  est le meilleur pour la norme 1 parmi ceux calculés et que celui de dimension  $2 \times 23$  est le meilleur

pour la norme 2. On peut maintenant comparer la construction améliorée avec ces motifs initiaux à la première construction.

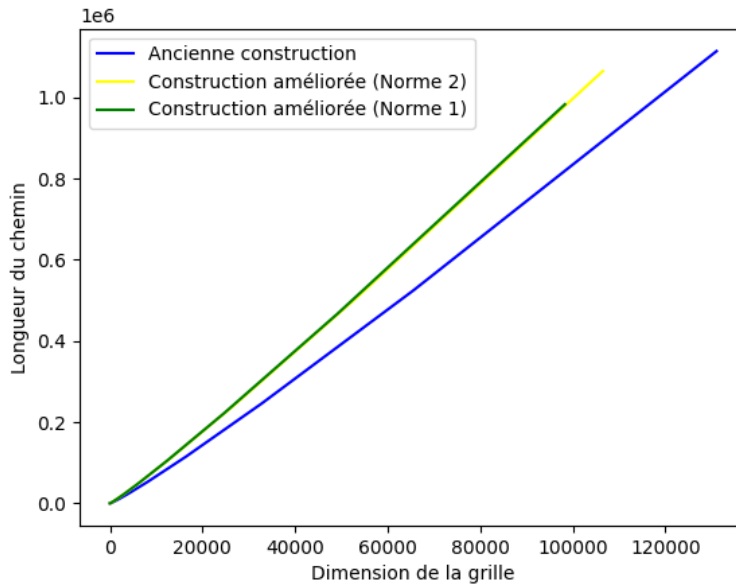


FIGURE 7 – Comparaison avec la construction précédente

On constate que le changement de motif initial a une influence minime par rapport à l’ajout d’un grand saut final.

On est ainsi parvenu à optimiser légèrement la méthode de construction. Après plusieurs tentatives de modification de cette méthode, on peut établir quelques conjectures. Il semble indispensable de se placer au centre en fin de motif pour permettre à la fois de laisser derrière des points non visités et de changer de direction. Il semble donc que la construction ne puisse pas être significativement améliorée puisque la longueur du chemin créé provient essentiellement des deux grands sauts du motif de construction.

### 2.2.2 Borne supérieure

On prouve ici une borne supérieure sur les grilles  $2 \times n$ .

**Lemme 2.1** *Considérons deux points sur une même colonne de la grille. La longueur du saut qui quitte le premier point parcouru est forcément 1.*

**Preuve.** Quand on parcourt le premier point de la colonne, le second n’est pas parcouru et il est à distance 1, donc le saut issu du premier point est majoré

par 1. Tous les sauts étants de longueur supérieure ou égale à 1, on en déduit que la longueur du saut en question est exactement 1.  $\square$

**Définition 2.6 (Couverture par sauts maximaux)** *Etant donné un chemin glouton dans une grille  $2 \times n$ , on s'intéresse à la projection selon l'axe horizontal des sauts qui constituent le chemin. On est alors ramené à un ensemble d'intervalles qu'on considérera ouverts. De plus, on ignore les intervalles vides créés.*

*Pour deux intervalles de l'ensemble fixés, il y a deux cas possibles : soit l'un est inclus dans l'autre, soit ils sont disjoints. On justifie l'exclusion du cas d'intersection non-vide après le lemme suivant cette définition.*

*Cette organisation permet d'envisager cet ensemble d'intervalles comme des arbres représentant la relation d'inclusion. Les sauts maximaux sont les racines de ces arbres, ils couvrent toute la longueur de la grille. Une majoration de la profondeur de ces arbres permettrait déjà d'obtenir une majoration de la longueur du chemin mais on peut faire plus fin.*

**Lemme 2.2** *Les points appartenant à une colonne couverte par l'intervalle associé à un saut sont parcourus avant ce saut.*

**Preuve.** Pour tous les points couverts sauf le plus éloigné du point de départ, ils sont toujours à une distance strictement inférieure à celle du saut donc sont visités avant le saut. Le point le plus éloigné est à distance égale à celle du saut avec la norme 1 si le saut est orthogonal (pour une norme  $p$  avec  $p > 1$  la distance est encore strictement inférieure).

Supposons par l'absurde qu'il ne soit pas visité avant le saut.

Cas 1 : Une seule colonne est couverte, le point visité de la colonne est entouré de points non visités, absurde.

Cas 2 : Considérons le point visité de la colonne couverte la plus éloignée. Soit c'est le point de départ du chemin (ne peut se produire qu'une fois), soit il est atteint après son voisin direct de la colonne précédente. On doit quitter ce sommet pour se placer au début du saut mais on est alors forcé de passer sur un de ses voisins qui doivent rester non visités (point d'arrivée du saut et point qu'on a supposé non visité, absurde.  $\square$

**Preuve** (Définition précédente)

Fixons deux intervalles  $]a, b[$ ,  $]c, d[$  avec  $a, b, c, d \in \mathbb{N}$ . Sans perte de généralité, on peut supposer  $a \leq c$ .

Supposons par l'absurde que ces intervalles ne sont pas disjoints et qu'aucun n'est inclus dans l'autre.

On a alors  $a \leq c < b < d$  donc  $a \leq c \leq b - 1 \leq d - 2$ . Aucun ordre de parcours des sauts associés aux intervalles ne convient puisqu'on devra toujours



quitter un point après qu'il ait été couvert, ce qui viole le lemme précédent.  $\square$

**Lemme 2.3** *Si une portion de la grille est encadrée par deux colonnes qui ont chacune au moins un point non visité, alors on peut majorer la longueur d'un saut partant d'un point de cette portion.*

*Pour une portion de dimension  $2 \times m$ , on peut majorer la composante horizontale du saut par  $\lfloor \frac{m+1}{2} \rfloor$  et la composante verticale par 1.*

**Preuve.** La longueur du saut est majorée par le minimum des distances aux points non visités connus (décroissance du minimum). Le pire cas est atteint si ces deux points sont alignés et que le départ du saut est sur la ligne opposée et sur une colonne centrée.  $\square$

**Lemme 2.4** *On peut majorer la longueur cumulée des sauts issus de sommets des  $m$  colonnes couvertes par un saut par la quantité  $\alpha(m)$  définie par :*

$$\alpha(0) = 0 ; \alpha(1) = 2 ; \alpha(m) = \max_{p \in \llbracket 1, \lfloor \frac{m+1}{2} \rfloor \rrbracket} p + 2 + \alpha(p-1) + \alpha(m-p)$$

**Preuve.** On prouve que  $\alpha(1) = 2$  convient par disjonction de cas (Fig. 13).

Cas général : On s'intéresse à la couverture par sauts maximaux de nos  $m$  colonnes. Considérons le saut  $S$  effectué en dernier parmi ceux issus de ces colonnes. Il mène forcément à la colonne du point de départ du saut couvrant. Si  $S$  est de longueur 1, on peut le supposer couvrant car il ne peut être couvert que par un saut de même longueur. Si  $S$  est plus long, il est forcément couvrant (lemme 2.2). On suppose que  $S$  est couvrant.

Notons  $p$  la longueur de la composante horizontale de  $S$ . On a  $p \in \llbracket 1, \lfloor \frac{m+1}{2} \rfloor \rrbracket$  d'après le lemme 2.3, applicable d'après le lemme 2.2. On peut majorer les sauts des colonnes couvertes par  $S$  par  $\alpha(p-1)$ . D'après le lemme 2.1, la longueur du saut issu du premier sommet parcouru sur la colonne du départ de  $S$  est 1.

Enfin les colonnes non traitées sont encadrées par deux points parcourus après qu'elles soient entièrement visitées : le point de départ de  $S$  et le point d'arrivée du saut couvrant. On a donc la même situation que s'il y avait un saut du point de départ de  $S$  au point d'arrivée du saut couvrant. Par conséquent, on peut majorer les sauts issus de ces colonnes par  $\alpha(m-p)$ .

La longueur de  $S$  est majorée par  $p+1$ . En sommant toutes les inégalités, on obtient que  $p+2+\alpha(p-1)+\alpha(m-p)$  majore la longueur cumulée des sauts issus des  $m$  colonnes pour  $p$  fixé. En prenant le maximum pour toutes les valeurs de  $p$  envisageables, on obtient un majorant indépendant de la structure du chemin sur ces  $m$  colonnes. Donc  $\alpha(m)$  est bien défini (noter que  $p-1 < m$  et  $m-p < m$  de sorte que la récurrence fonctionne correctement).  $\square$

**Lemme 2.5**  $\alpha(m) \leq 2m + \frac{m}{2} \log_2(1 + m)$ .

**Preuve** par récurrence forte.

Pour  $m < 3$ , l'inégalité se vérifie par le calcul.

Pour  $m \geq 3$ , soit  $p \in [1, \frac{m+1}{2}]$

On dispose des inégalités suivantes :

$$(1) : \log_2(p) \leq \log_2(m+1) - 1, (2) : -\frac{\log_2(m+1)}{2} \leq -1, (3) : \log_2\left(\frac{m+1-p}{m+1}\right) \leq -1$$

$$p + 2 + \alpha(p-1) + \alpha(p-m) \leq p + 2m + \frac{p-1}{2} \log_2(p) + \frac{m-p}{2} \log_2(m-p+1)$$

(hypothèse de récurrence)

$$p + 2 + \alpha(p-1) + \alpha(p-m) \leq 2m + p + \frac{p-1}{2}(\log_2(m+1) - 1) + \frac{m-p}{2} \log_2(m-p+1)$$

par (1)

$$p + 2 + \alpha(p-1) + \alpha(p-m) \leq 2m + \frac{m}{2} \log_2(1+m) + \frac{p+1}{2} + \frac{m-p}{2} \log_2\left(\frac{m+1-p}{m+1}\right) - \frac{\log_2(m+1)}{2}$$

$$p + 2 + \alpha(p-1) + \alpha(p-m) \leq 2m + \frac{m}{2} \log_2(1+m) + \frac{p+1}{2} - \frac{m-p}{2} - 1$$

par (2) et (3)

$$p + 2 + \alpha(p-1) + \alpha(p-m) \leq 2m + \frac{m}{2} \log_2(1+m) + p - \frac{m+1}{2}$$

$$p + 2 + \alpha(p-1) + \alpha(p-m) \leq 2m + \frac{m}{2} \log_2(1+m)$$

$$\text{donc } \alpha(m) \leq 2m + \frac{m}{2} \log_2(1+m) \quad \square$$

Pour obtenir un majorant du chemin entier, on applique la majoration  $\alpha$  sur les sauts maximaux couvrants.

**Lemme 2.6** *En notant  $l_i$  les longueurs des composantes horizontales des sauts maximaux associés à un chemin glouton, on peut majorer la longueur de ce chemin par  $1 + \sum_i (l_i + 2 + \alpha(l_i - 1))$*

**Preuve.** Considérons la colonne du dernier point parcouru, elle se trouve forcément en extrémité d'un saut maximal (lemme 2.2). De plus, le seul saut issu de cette colonne est de longueur 1 (lemme 2.1). Si ce saut est un saut maximal, on est certain d'arriver au dernier point par le côté vers lequel ce saut est orienté. En effet, il est impossible de dépasser le point final sans le visiter. Et comme le saut considéré est maximal, on arrive au point d'arrivée par un saut de longueur 1 dont l'intervalle associé est le même que celui de notre saut maximal. On peut donc sans perte de généralité considérer que le saut maximal est celui qui arrive au point final.

On peut donc toujours considérer que la colonne du point final est une extrémité de saut maximal de laquelle ne part pas de saut maximal. Les longueurs issues de la colonne dont part le  $i$ -ème saut maximal peuvent être majorées par  $(l_i + 1) + 1$ . Les longueurs issues des colonnes couvertes par ce saut sont majorées par  $\alpha(l_i - 1)$ .

On a considéré toutes les colonnes et, en sommant les majorations, on

obtient  $1 + \sum_i (l_i + 2 + \alpha(l_i - 1))$ . □

**Théorème 2.1** *Un chemin glouton dans une grille  $2 \times n$  est de longueur au plus  $3n + \frac{n}{2} \log_2(n) - 2$*

**Preuve.** Notons  $k$  le nombre de sauts maximaux. La longueur cumulée des sauts maximaux est  $n - 1$  et  $x \mapsto \frac{x}{2} \log_2(1 + x)$  est sur-additive donc :

$$1 + \sum_i (l_i + 2 + \alpha(l_i - 1)) \leq 1 + (n - 1) + 2k + 2(n - k - 1) + \frac{n - k - 1}{2} \log_2(n - k) \leq -2 + 3n + \frac{n}{2} \log_2(n) \quad \square$$

Pour le cas orthogonal, on n'a pas à majorer les composantes verticales des sauts. On obtient donc une majoration plus fine en adaptant les preuves précédentes.

**Corollaire 2.1** *Un chemin glouton orthogonal dans une grille  $2 \times n$  est de longueur au plus  $2n + \frac{n}{2} \log_2(n) - 1$*

### 2.3 Grilles $k \times n$

On adapte aisément le motif utilisé pour les grilles  $2 \times n$  en prolongeant les déplacements verticaux.

**Proposition 2.2** *Pour une grille de dimension  $k \times 2^{i+3}$ , on peut construire un chemin glouton orthogonal de longueur  $2^{i+2}(2k + i + 2) - 1$ .*

*On montre ainsi que le ratio d'approximation est au moins de  $\frac{\log_2(n)}{2k}$  pour une grille  $k \times n$ .*

La preuve de borne supérieure ne s'adapte pas aussi immédiatement mais cela reste envisageable. L'intuition suggère que le ratio sera plus petit quand on augmente le nombre de lignes. En effet, dans une grille  $2 \times n$ , on pouvait effectuer un grand saut après avoir visité un nombre de points linéaire en la taille du saut, mais cela ne sera plus toujours vrai si on augmente le nombre de lignes. Pour un saut de longueur de l'ordre de  $k$ , il faudra avoir visité un nombre quadratique de points.

Les calculs de plus longs chemins sur de petites grilles semblent confirmer cette intuition.

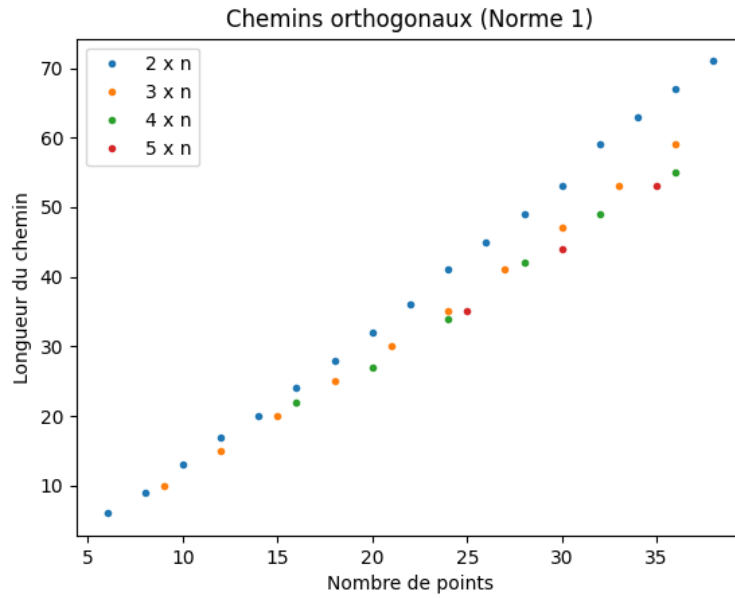


FIGURE 8 – Plus longs chemins gloutons orthogonaux sur des grilles de dimension variable

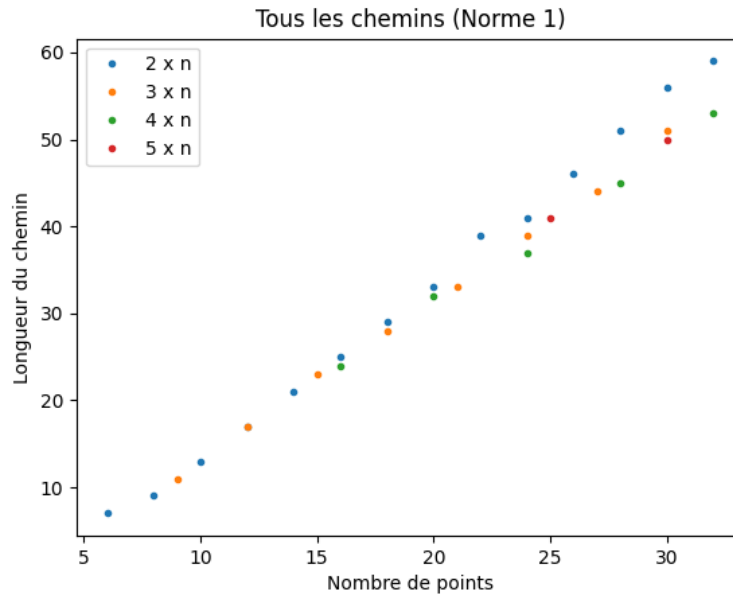


FIGURE 9 – Plus longs chemins gloutons sur des grilles de dimension variable

## 2.4 Conclusions

On a obtenu un encadrement plus précis du plus long chemin glouton dans les grilles  $2 \times n$ , comme le montre le graphique suivant :

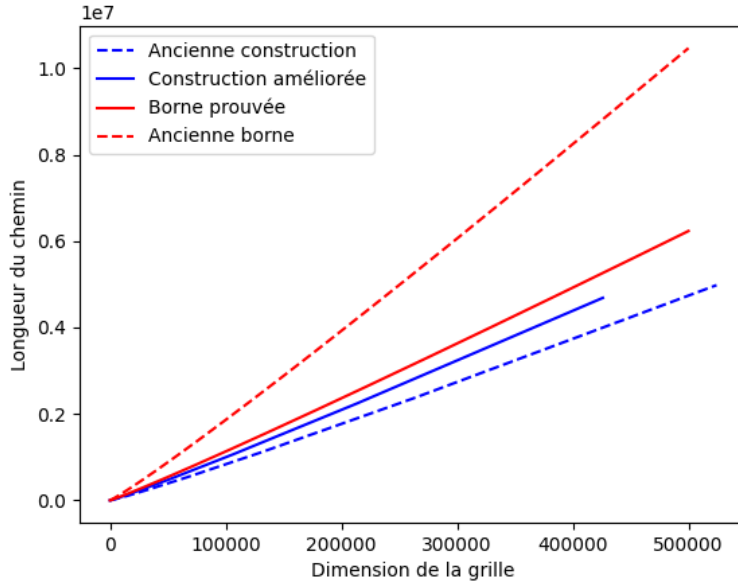


FIGURE 10 – Encadrement obtenu

On peut conclure que le ratio sur les grilles  $2 \times n$  est  $\frac{1}{4} \log_2(n) + o(\log_2(n))$ .

On imagine que les grilles de dimensions supérieures ne permettront pas d'obtenir un ratio d'approximation plus grand sans modification. On peut toutefois envisager des grilles fractales qui permettent d'augmenter plus rapidement la longueur des sauts.

## A Annexe

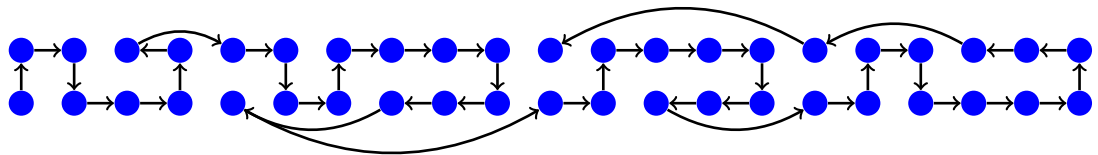


FIGURE 11 – Motif  $2 \times 21$  pour la norme 1

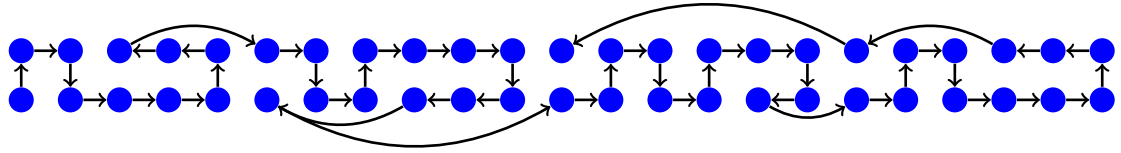


FIGURE 12 – Motif  $2 \times 23$  pour la norme 2

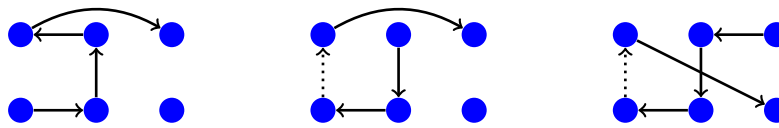


FIGURE 13 – Cas possibles pour la couverture d'une colonne

## Références

- [BH15] Judith Brecklinghaus and Stefan Hougardy. The approximation ratio of the greedy algorithm for the metric traveling salesman problem. *Operations Research Letters*, 43(3) :259–261, May 2015.
- [HW15] Stefan Hougardy and Mirko Wilde. On the nearest neighbor rule for the metric traveling salesman problem. *Discrete Mathematics*, 195 :101–103, November 2015.
- [Pap77] Christos H. Papadimitriou. The euclidean travelling salesman problem is np-complete. *Theoretical Computer Science*, 4(3) :237,244, June 1977.
- [RSLI77] Daniel J. Rosenkrantz, Richard E. Stearns, and Philip M. Lewis II. An analysis of several heuristics for the traveling salesman problem. *SIAM Journal on Computing*, 6(3) :563–581, September 1977.