SAÉ Optimisation 2025-2026

Autour du tracé d'un métro circulaire

On s'intéresse au problème du tracé d'une ligne de transport public circulaire (cela peut être un bus, un tramway ou un RER: toute ressemblance avec les métros du grand Paris serait parfaitement fortuite). Pour cela nous utiliserons des données réalistes disponibles en ligne, et qui correspondent à de vraies localisations, pour un problème proche qui est celui du voyageur du commerce.

Les instances du problème considéré ici seront composées d'un ensemble de points sur une carte correspondant à des centres de zones à fort peuplement où peuvent être installées potentiellement des stations de cette ligne. On souhaite construire le tracé de la ligne passant par seulement p de ces points: les usagers des zones alentour iront à pied rejoindre la station la plus proche. L'objectif est de minimiser la longueur des trajets à pied et en transport pour les usagers de la ligne.

1 La problématique des transports publics

Il est à noter que d'autres paramètres et d'autres objectifs pourraient être choisis avec pertinence pour le tracé d'une ligne de transports publics. En effet, les agglomérations réfléchissent régulièrement à construire ou reconstruire les services de transports publics de manière à pouvoir ensuite estimer les coûts engendrés par cette construction et tenter également d'en déduire dans un second temps les coûts d'exploitation. Il s'agit donc de tracer des lignes de transports (bus, métro, tramway...), de fixer le nombre de véhicules sur les lignes, les horaires de ces lignes, l'affectation des conducteurs etc.

La conception d'un plan de transport pour une agglomération se doit de respecter certains principes, parmi lesquelles:

- couverture: elles doivent couvrir au mieux l'agglomération.
- connexité: elles doivent proposer un ensemble connexe, permettant d'aller d'un point à l'autre du réseau.
- interconnection: elles doivent posséder de nombreuses interconnexions, c'est-à-dire des points où deux (ou plus) lignes se croisent. D'autre part, les horaires doivent être compatibles pour éviter les attentes.
- équité: elles doivent être plus nombreuses et fréquentes dans les zones de forts passages (lieux de vie, travail).
- lisibilité: les lignes doivent rester lisibles pour les usagers: pas trop courtes, pas trop longues, transversales pour l'agglomération...

Le problème considéré ici n'est donc qu'un aspect réduit de cette vaste problématique.

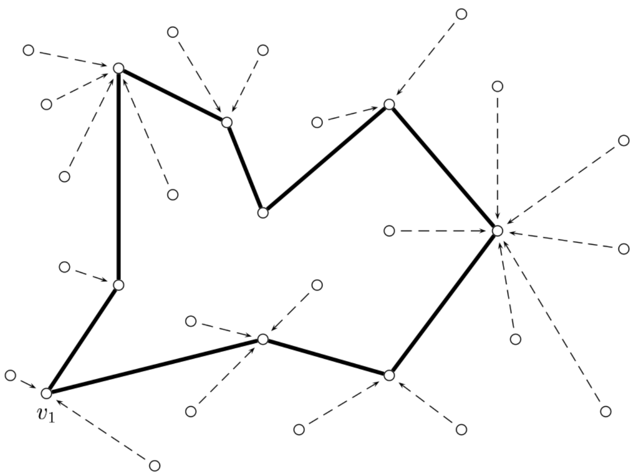
2 Formalisation

On considère n points, rassemblés dans un ensemble V, dans une agglomération où chaque point est repéré par des coordonnées latitude-longitude $(l_i, L_i), i \in \{1, ..., n\}$. Ces points représentent des zones bien délimitées de l'agglomération (quartiers, zones industrielles,...). Le calcul de la distance euclidienne $d_{ij} = \sqrt{(l_i - l_j)^2 + (L_i - L_j)^2)}$ entre deux points $i, j \in \{1, ..., n\}$ fournit une estimation de la longueur d'un trajet à pied ou en transport entre i et j. On considère le graphe complet K_n induit par les n points. La ligne que l'on désire tracer passera par certains des points qui seront alors appelés *stations*. Notez que pour ce problème, $n \geq p \geq 3$.

Un anneau-étoile (Ring-Star, en anglais) est défini par :

- p points parmi les n points qui seront les stations,
- une affectation de chacun des points qui ne sont pas stations à l'une des stations,
- un ensemble d'arêtes reliant les p stations en formant un cycle.

L'illustration suivante est empruntée à l'article [1]. Sur cette illustration les stations sont les points où passent le métro circulaire représenté par des arêtes en gras. Les affectations d'un point à une station (qui sont les déplacements à pieds des usagers vers les stations) sont représentées par des arcs en pointillés.



Le *problème de l'anneau-étoile* (the Ring-Star problem, en anglais) consiste à déterminer un anneau-étoile tel que la somme des distances euclidiennes des arêtes de l'anneau plus des distances des arcs des affectations soient minimales. Compte-tenu de cette définition, il est à noter que les choix d'affectations des points aux stations se feront toujours à la station la plus proche. **Dans ce projet, on considére que le sommet 1 est toujours une station.**

Ce problème est ainsi mono-critère car on somme toutes les distances. Il est à noter que la nature des longueurs des arêtes (distance inter-stations) et celle des arcs (longueur des trajets à pied) ne sont peut-être pas à mettre sur un même plan. Il est envisageable de prendre en considération un paramètre $0 \leq \alpha \leq 10$ multiplicateur des distances inter-stations pour témoigner de l'importance relative des deux termes de la somme.

Remarque: Dans le cas où $p = n$, le problème de l'anneau-étoile est le problème du voyageur de commerce.
Dans le cas où $\alpha = 0$, le problème est celui du p-médian.

3 Instances

Nous ne disposons pas d'instances directement adaptées à ce problème, mais nous allons utiliser les instances provenant d'une librairie d'instances réalistes dédiées au problème du voyageur de commerce, la TSPlib: `http://comopt.ifi.uni-heidelberg.de/software/TSPLIB95`.

Dans cette librairie d'instances, nous nous intéressons aux instances symétriques et parmi elles aux instances données simplement par un nuage de points et leurs coordonnées, comme par exemple att48 qui correspond aux 48 capitales des états connexes des Etats-Unis (on ne considère pas les instances qui sont d'un autre format).

La visualisation de ces instances se limite alors à celle du nuage de points. En revanche, une visualisation des solutions permet de visualiser le tracé et les arcs d'affectation.

Remarque importante: On va considérer tout au long de la SAÉ que le premier sommet (numéroté 0 ou 1) sera toujours une station.

4 Etat de l'art

Ce problème de conception d'une ligne de métro circulaire peut être vu comme un problème d'optimisation combinatoire rassemblant deux problèmes classiques:

- le problème du p-médian pour la détermination des stations,
- le problème du voyageur de commerce pour la détermination de l'ordre de visite des stations.

On appelle parfois ce type de problème presque classique un problème "combiné" ou "intégré".

4.1 Problème du p-médian

Le problème du p-median a pour instance un ensemble de n points rassemblés dans V et des distances d_{ij} entre les points $i, j \in V \times V$. Il consiste à sélectionner, parmi les n points, p points qui seront appelés les médians (ou centres); ainsi que déterminer des affectations des points non médians aux médians de manière à minimiser la somme totale des distances des affectations.
Il est à noter que si les distances d_{ij} sont euclidiennes, les affectations des points non-médians aux médians consiste à affecter un point non-médian au médian le plus proche.

Résolution heuristique du problème du p-médian

▶ *Heuristique gloutonne:*
Voici une idée pour une méthode gloutonne produisant une solution de bonne valeur en cas de points répartis assez uniformément.

- On repère les dimensions max et min des points sur les 2 axes du plan.
- On pose $q = \lceil \sqrt{p} \rceil$ et on divise le plan en $q \times q$ rectangles de même taille.
- Pour chaque rectangle, on liste tous les points contenus dans le rectangle et on en cherche le point le plus proche du centre du rectangle.
- On regroupe des paires de points proches pour avoir exactement p médians.
- Affecter les points non-médians au médian le plus proche.

Attention, si les points sont mal répartis, des rectangles peuvent être vides et l'heuristique produit moins de p médians.

▶ *Heuristique randomisée*

Une autre heuristique plus simple mais pouvant produire une très mauvaise valeur est la suivante:

- Tirer au sort p points parmi les n points.
- Affecter les points non-médians au médian le plus proche.

Il est possible de réitérer cette heuristique un certain nombre de fois et de conserver le meilleur tirage.

Notez qu'il est possible de cumuler les deux heuristiques précédentes pour être certain de produire toujours une solution.

▶ *Métaheuristique itérative*

Il est très utile d'améliorer une solution heuristique en utilisant une méta-heuristique. Dans le cas du problème du p-médian avec distance euclidienne, on peut noter qu'un encodage possible est de considérer un ensemble de p médians (par exemple codé à la fois comme un vecteur binaire sur les n points, plus une liste chaînée des p médians). Avec un tel encodage, une méthode itérative (descente stochastique itérée, recuit simulé, tabou) peut être facilement implémentée sur la base d'un voisinage qui échange un point médian et un point non médian.

Formulation PLNE compacte pour le problème du p-médian:

On considère les variables binaires y_{ij} pour tout couple i, j de points dans V:

- y_{ii} vaut 1 si i est un median et 0 sinon
- y_{ij} vaut 1 si le point j est un médian, i n'est pas un médian et i est affecté à j.

Le programme linéaire en nombres entiers suivant est équivalent au problème du p-médian:

$$(P)\left\{\begin{array}{llll} Min & \displaystyle\sum_{(i,j)\in V\times V} d_{ij}y_{ij} & & \\ & \displaystyle\sum_{i\in V} y_{ii} = p & & (1) \\ & \displaystyle\sum_{j\in V} y_{ij} = 1 & \forall i \in V & (2) \\ & y_{ij} \leq y_{jj} & \forall (i,j) \in V \setminus \{j\} \times V & (3) \\ & y_{ij} \in \{0,1\} & \forall (i,j) \in V \times V & \end{array}\right.$$

Cette équivalence peut être prouvée par les arguments suivants. Considérons une solution entière y de (P): les variables y_{ii} donnent par l'inégalité (1) p points que l'on peut considérer comme les p médians. Par (3), si un point n'est pas médian alors aucun point ne lui est affecté. Enfin, par (2) on peut voir que si un point est médian, il n'est pas affecté à un autre point; et s'il n'est pas médian, il est affecté à exactement un point qui par (1) n'est autre que lui-même et qui par (3) ne peut donc qu'être un médian. Donc y décrit bien une solution du p-médian. Inversement, toute solution du p-médian est clairement solution de (P) et les fonctions objectifs coïncident.

Cette formulation est compacte avec $O(n^2)$ variables binaires et $O(n^2)$ contraintes. Elle peut être confiée à un solveur entier pour être résolue. La littérature rapporte que les solveurs sont capables de résoudre des tailles d'instances assez importantes (autour de la centaine de sommets).

4.2 Problème du voyageur de commerce symétrique

Ce problème a été vu en cours.

Résoudre heuristiquement le problème du voyageur de commerce par une heuristique

Implémenter l'heurisique du plus proche voisin qui consiste à partir du sommet 1 à choisir itérativement parmi les autres points celui qui est le plus proche.
Cet algorithme doit être amélioré par une méthode itérative mettant en œuvre le voisinage 2-OPT.

Formulations PLNE compactes et non-compactes pour le problème du voyageur de commerce:

Plusieurs formulations pour le problème du voyageur de commerce (symétrique ou assymétrique) sont possibles, notamment:
- 3 formulations compactes: une basée sur les contraintes MTZ, une dite de flot agrégé et une dite de flot désagrégé.
- 1 formulation à nombre exponentiel d'inégalités (version assymétrique et symétrique)

La formulation MTZ est battue par la formulation de flot agrégé. La formulation non-compacte exponentielle est beaucoup plus efficace que les deux premières formulations compactes. La formulation à flot désagrégé, bien que compacte, contient un nombre très important d'inégalités qui bloquent sa résolution. On privilégie alors plutôt la formulation à flot agrégé si l'on désire une formulation compacte, ou sinon la formulation non-compacte.

5 Formulation pour le problème de l'anneau-étoile

Pour rappel, dans la version étudiée, le sommet 1 est toujours une station.

5.1 Formulation compacte

On s'intéresse à la formulation compacte suivante.
Les variables y sont similaires aux variables d'affectation du p-médian: les valeurs $y_{ii} = 1$ désignent les p stations.
Les variables x sont similaires aux variables déterminant les arêtes à prendre pour le problème du voyageur de commerce: on voit que tous les sommets passant par une station seront de degré 2: ainsi toute station appartient à une cycle.
Les variables z sont des variables, dites de flots, où un flot de valeur $p-1$ part du sommet 1 et se trouve décrémenter à chaque passage par une station (repérée par un sommet où $y_{ii} = 1$). Le flot ne circule que sur des arcs où $x_{ij} = 1$. Les contraintes de flot empêchent l'existence d'un cycle qui ne passe pas par le sommet 1.

$$
\begin{array}{lll}
Min \sum_{ij \in E} d_{ij} x_{ij} + \sum_{(i,j) \in V \times V} d_{ij} y_{ij} & & \\
\sum_{i \in V} y_{ii} = p & & (1) \\
\sum_{j \in V} y_{ij} = 1 & \forall i \in V & (2) \\
y_{ij} \leq y_{jj} & \forall (i,j) \in V \setminus \{j\} \times V & (3) \\
\sum_{ij \in \delta(i)} x_{ij} = 2y_{ii} & \forall i \in V & (4) \\
\sum_{j \in V \setminus \{1\}} z_{1j} = p - 1 & & (5) \\
\sum_{j \in V \setminus \{i\}} z_{ji} = \sum_{j \in V \setminus \{1,i\}} z_{ij} + y_{ii} & \forall i \in V \setminus \{1\}, & (6) \\
z_{ij} + z_{ji} \leq (p-1) x_{ij} & \forall i \in V, j \in V \setminus \{1, i\} & (7) \\
x_{ij} \in \{0,1\} & \forall ij \in E & \\
y_{ij} \in \{0,1\} & \forall (i,j) \in V \times V & \\
z_{ij} \in [0, p-1] & \forall (i,j) \in V \times V \setminus \{1, j\} & \\
y_{11} = 1 & & \\
y_{1j} = 0 & \forall j \in V \setminus \{1\} &
\end{array}
$$

5.2 Formulation non-compacte

Dans cette formulation, on retrouve les variables x et y précédentes.
L'inégalité (4) force la connexité entre le sommet 1 et tout sommet j qui n'est pas un médian.

Une première formulation

$$
\begin{array}{lll}
Min \sum_{ij \in E} d_{ij} x_{ij} + \sum_{(i,j) \in V \times V} d_{ij} y_{ij} & & \\
\sum_{i \in V} y_{ii} = p & & (1) \\
\sum_{j \in V} y_{ij} = 1 & \forall i \in V & (2) \\
y_{ij} \leq y_{jj} & \forall (i,j) \in V \setminus \{j\} \times V & (3) \\
\sum_{ij \in \delta(i)} x_{ij} = 2y_{ii} & \forall i \in V & (4) \\
\sum_{e \in \delta(S)} x(e) \geq 2y_{ii} & \forall S \subset V, 1 \notin S, i \in S & (8) \\
x_{ij} \in \{0,1\} & \forall ij \in E & \\
y_{ij} \in \{0,1\} & \forall (i,j) \in V \times V & \\
y_{11} = 1 & & \\
y_{1j} = 0 & \forall j \in V \setminus \{1\} &
\end{array}
$$

Les inégalités (8) étant en nombre exponentiel, il est nécessaire de réfléchir à des algorithmes de séparation.

Ajouts de certaines inégalités de manière compacte

On peut remarquer que les inégalités (8) pour $|S| = 2$, donc $S = \{i, j\}$ avec $i \neq 1$ s'écrivent également $2y_{jj} \geq x_{ij}$ qui est elle-même dominée par

$$y_{jj} \geq x_{ij} \quad \forall i \in V \setminus \{1\}, j \in V (9)$$

On peut donc ajouter les inégalités (9) dès le départ parmi les inégalités "en dur".

Séparation entière

Etant donné un vecteur (x, y) entier, la séparation des inégalités (8) revient à:
- repérer les p points médians (tels que $y_{ii} = 1$),
- si les p points sont reliés par un unique cycle formé par les arêtes telles que $x_{ij} = 1$: alors (x, y) vérifient toutes les inégalités (8),
- sinon les sommets formant un cycle ne passant pas par 1 forment un ensemble S correspondant à une inégalité violée (8).

Séparation fractionnaire

Voici la **séparation exacte** des inégalités (8) quand (x, y) est fractionnaire.
Etant donné un vecteur (x, y) fractionnaire, la séparation des inégalités (8) revient à:
- Rechercher l'arbre de Gomory-Hu pour les arêtes valuées par x,
- Pour tout sommet i différent de 1 :
  Rechercher si la coupe minimale C séparant i et 1 est inférieure à $2y_{ii} - \epsilon$, $\epsilon = 0.001$
  Dans ce cas, l'ensemble S contenant i qui engendre C correspond à une inégalite (8) violée
Si aucune inégalité n'a été produite, toutes les inégalités (8) sont vérifiées par (x, y).

La recherche de l'arbre de Gomory-Hu est donnée par une fonction de l'outil Lemon de complexité $O(n^4)$.
L'arbre de Gomory-Hu est un arbre sur l'ensemble des sommets du graphe, mais il peut contenir des arêtes qui ne sont pas dans le graphe. Il a la propriété que la coupe de capacité minimale séparant deux sommets i et j dans le graphe est égale à la valeur minimale portée par une arête sur le chemin de l'arbre reliant i et j. De plus, en ôtant cette arête dans l'arbre, on obtient une partition des sommets en deux ensembles S_i (resp. S_j) contenant i (resp. j) telle que la coupe sortant de S_i (qui est la même que celle sortant de S_j) dans le graphe est exactement la coupe minimale séparant i et j.

Il est à noter que l'on peut, dans le cas où la séparation entière a été implémentée, se "contenter" d'une **séparation heuristique**. Une séparation heuristique consiste alors à rechercher heuristiquement un ensemble S qui correspond à une inégalité (8) violée: cela peut se faire de manière gloutonne. Par exemple, pour tout sommet $i \neq 1$ de valeur $y_{ii} > 0$, on part de $S = \{i\}$ et on ajoute peu à peu des sommets dans S tant que l'inégalité n'est pas violée. On peut noter que cette heuristique ne répond que partiellement au problème de séparation car il peut échouer sans prouver qu'il reste des inégalités non violées. En revanche, il est bien plus rapide que la séparation exacte.

Mise en œuvre
Il est à noter que la séparation entière est suffisante pour rendre valide la formulation.
Ceci rend donc possible l'utilisation de la séparation heuristique pour le cas fractionnaire.

La séparation exacte du cas fractionnaire n'est à utiliser que si la version heuristique échoue.

Une deuxième formulation

Il a été prouvé par l'étude du polyèdre de l'anneau réalisé dans [1] que les inégalités (3) et (8) pouvaient être remplacées par les inégalités suivantes.

$$\sum_{e\in\delta(W)} x(e) \geq 2\sum_{j\in W} y_{ij} \quad \forall W \subset V, 1 \notin W, i \in W \qquad (10)$$

Les inégalités (10) définissent, sous certaines conditions, des facettes du polyèdre: elles sont ainsi plus fortes pour renforcer la valeur de relaxation et amener à davantage de cas où les relaxations linéaires fournissent des solutions entières. De plus, la séparation des inégalités (10) est plus efficace en terme de complexité que celles des inégalités (8) grâce à une astuce de réécriture: cette procédure est également décrite dans [1].

6 Travail démandé

Il est demandé de faire, seul ou en binôme, tous les points suivants:
- montrer la NP-difficulté du problème,
- une visualisation des données et des solutions,
- l'implémentation d'une heuristique et d'une métaheuristique,
- l'implémentation d'une formulation compacte,
- comparer les résultats obtenus pour les 3 méthodes: temps de calcul et durée d'exécution.
- un rendu sous forme d'un mini-rapport et des archives de votre code.
- une soutenance de vos travaux avec mise en pratique sur machine.

En Bonus, vous pourrez également faire l'implémentation d'une formulation non-compacte,

6.1 NP-difficulté

Montrez, en utilisant les informations données dans ce sujet, que le problème de l'anneau-étoile est NP-difficile.

6.2 Entrée-Sorties et visualisation

Faites une visualisation graphique pour les instances et vos solutions.

Aide:
- le code proposé pour le TSP sur moodle propose une lecture des instances du TSP, ainsi qu'une formulation compacte et des heuristiques.
- il contient également une visualisation assez simple que vous pouvez adapter à ce problème.

6.3 Heuristique

Une première démarche très naturelle est d'enchaîner la résolution des deux problèmes du p-médian et du voyageur de commerce. Il s'agit bien d'une méthode heuristique car il faut résoudre ces deux aspects du problème simultanément pour obtenir une solution optimale.

Rappel: Le sommet 1 (ou 0 s'il existe) des instances est toujours un médian dans l'énoncé du problème.

Une façon de faire cet enchaînement serait d'utiliser la PLNE pour le problème du p-médian: mais cette solution reste limitée à une taille réduite à quelques centaines de sommets.

Enchaîner les heuristiques est aussi possible pour résoudre des instances de grandes tailles (au prix de la qualité bien-entendu).

Proposer également une métaheuristique améliorant itérativement la solution produite et qui prend en compte globalement les voisinages du p-médian et du TSP tout en visant à améliorer le coût total de l'anneau-étoile.

6.4 Formulation compacte

Mettre en œuvre la formulation compacte.
Noter que l'ajout des inégalités (9) peut aussi être réalisée dans ce contexte.

[**Bonus**] Proposer d'autres formulations compactes en vous basant sur la littérature scientifique.

6.5 [Bonus] Formulation non-compacte

Mettre en œuvre la **première** formulation non-compacte proposée:
- en ajoutant les inégalités (9) en dur
- en implémentant la séparation des inégalités (8) dans le cas entier
- en implémentant la séparation heuristique des inégalités (8) dans le cas fractionnaire.

Améliorer la formulation non-compacte en tenant compte de la séparation exacte des inégalités (8) ou en considérant la deuxième formulation basée sur les inégalités (10) ainsi que la littérature scientifique.

6.6 Jeux d'essais et comparaisons numériques

Comparer les diffférentes méthodes sur un lot suffisant d'instances de tailles croissantes. Penser également à faire varier p.

6.7 Rapport et code

Rédiger un rapport où figurera notamment les réponses aux différentes questions, vos explications, les algorithmes (mais pas les codes) proposés, les résultats expérimentaux et leur comparaison. Le rapport sera déposé au format pdf sur Moodle avec le code dans une archive tar.gz, tgz ou zip le vendredi 30 janvier 2025 au plus tard. Pour le code (dans le langage que vous pouvez choisir), celui-ci devra être fourni avec un Makefile et un README pour son installation, sa compilation et son utilisation et il devra tourner sous Linux, par défaut et a minima sur les machines des salles TP de l'Institut Galilée. Une soutenance sera organisée avec mise en pratique sur machine.

References

[1] M. Labbé, G. Laporte, I.R. Martin et J.J. S. González. The ring star problem: Polyhedral analysis and exact algorithm. *Networks* 43:3, 177–189 (2004).