Projet Vélib
Équilibrage nocturne d'un parc de vélos

Enseignant responsable : Éric Soutil (eric.soutil@cnam.fr)

1) Objectif

L'objectif de ce projet est l'étude d'un problème d'optimisation du domaine environnemental. Il s'agira de modéliser puis de résoudre un problème lié à la gestion d'un parc de vélos en libre service. Plusieurs modélisations de ce problème sont envisageables. On suggère ici une modélisation quadratique en variables 0-1 et entières, mais toute autre modélisation (correcte) sera acceptée, en particulier une modélisation linéaire plus classique inspirée du problème de voyageur de commerce (avec capacité). Le projet sera l'occasion mettre en œuvre une méta-heuristique.

2) Description du projet

Une entreprise gère un parc de vélos en libre service dans une grande ville. Un grand nombre de stations sont installées dans la ville. Un usager peut prendre un vélo disponible dans une station A et se rendre à une station B où il dépose le vélo qu'il a emprunté en A (l'emprunt du vélo est gratuit durant une durée limitée à 30 minutes, des bornes vérifient la durée d'emprunt à chaque station). Certains endroits de la ville étant plus attractifs que d'autres, le nombre de vélos disponibles à chaque station fluctue énormément au cours de la journée. Cela pose un double problème : un usager peut ne trouver aucun vélo disponible dans une station « peu attractive » (ou simplement située en hauteur comme certaines stations du 20ème arrondissement à Paris), ce qui entraîne un mécontentement certain ; a contrario, aux endroits très touristiques, il se peut qu'une station soit saturée (plus aucun emplacement n'est libre pour déposer son vélo), obligeant le cycliste à se rendre à une autre station non saturée, ce qui est également gênant. Pour pallier ces inconvénients, l'entreprise gérant le parc de vélos équilibre la nuit les stations : un semi-remorque circule entre les stations, désengorge les stations saturées et au contraire dépose des vélos aux stations en sous-effectif.

Le but de ce projet est la prévision de la tournée du semi-remorque qui équilibre les stations de vélos.

- **Hypothèses et contraintes à respecter** : la tournée du semi-remorque commence au magasin, où est gardé un stock suffisamment grand de vélos. On charge au magasin le semi-remorque avec un certain nombre de vélos, ce nombre fait partie des inconnues du problème. Puis le semi-remorque doit passer par chacune des n stations qui composent le parc, y compris par les stations déjà équilibrées. En effet, cette tournée est également l'occasion de contrôler l'état des stations afin de programmer d'éventuelles interventions. Si on ne faisait pas cette hypothèse, un pré-traitement supprimant de la tournée les stations déjà équilibrées serait immédiat. La tournée se termine par le retour du semi-remorque au magasin, pas nécessairement à vide. Chaque station n'est visitée qu'une unique fois. On doit naturellement respecter la capacité du semi-remorque.

- **Données du problème**. On connaît à l'avance :
  - Le nombre $\boldsymbol{n}$ de stations concernées par la tournée à organiser et leur emplacement géographique : les stations sont disposées dans un repère dont l'unité est l'hectomètre (1u=100m), à l'intérieur d'un carré de 10km de côté. Concrètement on connaît les coordonnées (x, y) de chaque station. Les coordonnées du magasin sont également connues ;
  - La capacité $\boldsymbol{K}$ du semi-remorque (c'est le nombre maximal de vélos qu'il peut transporter) ;
  - La distance $\boldsymbol{d_{ij}}$ entre deux stations i et j ainsi que la distance du magasin à chacune des stations se déduira directement des coordonnées des stations et du magasin, en arrondissant obligatoirement la distance euclidienne calculée à l'entier le plus proche (méthode round).

- Le nombre $nbvp_i$ de vélos effectivement présents à la station i ainsi que la capacité maximale cap_i de la station i;
- Le nombre $ideal_i$ qui représente le nombre idéal (ni trop ni trop peu) de vélos qui doivent se trouver à la station i pour permettre un bon fonctionnement de la station (ce nombre est de l'ordre de 75% du nombre total de places de la station, mais il dépend également de l'emplacement des stations, il est quoi qu'il en soit connu).
- Remarque : des vols pouvant survenir, on a : $\sum_{i=1}^{n} nbvp_i \leq \sum_{i=1}^{n} ideal_i$

- **Objectifs visés pour la tournée** : on se pose deux questions. La première est de savoir si, connaissant les données du problème, il est ou non possible de remettre chaque station à son niveau idéal de vélos en une unique tournée. On dira qu'une station a été équilibrée lorsqu'elle a été remise à son nombre idéal de vélos. Si, après la tournée, une station reste en manque de vélos par rapport à son idéal, le nombre de vélos manquant sera appelé « déséquilibre de la station » (ce nombre est positif ou nul). De même, si après la tournée une station a un excédent de vélos par rapport à son idéal, cet excédent sera également appelé déséquilibre de la station (ce nombre est également positif ou nul). **Un premier objectif** consiste donc à organiser la tournée de sorte à minimiser (et si possible annuler) le déséquilibre global, défini comme la somme des déséquilibres de chaque station. La deuxième question qui nous intéresse est de savoir parmi l'ensemble des tournées possibles qui minimise le déséquilibre global, laquelle permet de parcourir la plus petite distance globale. Un **second objectif**, moins prioritaire que le premier consiste donc à minimiser la distance totale parcourue après avoir minimisé le déséquilibre, pour des raisons évidentes d'efficacité et d'économie d'énergie. En ce sens, une **solution optimale** de notre problème sera donc une solution pour laquelle la valeur d^* du déséquilibre global est minimale et pour laquelle la distance totale parcourue est la plus petite parmi toutes les solutions ayant une valeur de déséquilibre égale à d^*.

- **Suggestions de modélisation**

  Plusieurs modélisations de ce problème peuvent être envisagées. Nous vous en suggérons une, quadratique, en variables 0-1 et entières dont le modèle mathématique sera à rendre le 12/04/2025. Une correction du modèle sous la forme d'un programme mathématique quadratique en variable 0-1 et entières et sa version linéarisée vous sera proposée le 13/04/2025.

  Dans la modélisation du problème suggérée, deux types de variables sont utilisées :

  - **des variables 0-1 x_{ij}** ayant la signification suivante : x_{ij} vaut 1 si et seulement si la station i est la $j^{\text{ème}}$ station visitée. La tournée effectue n « étapes » (une étape est la visite d'une station). Le départ du magasin et le retour au magasin, tous deux imposés, ne sont pas considérés comme des étapes du problème (mais les distances parcourues du magasin à la station de la première étape et celle de la station de la dernière étape au magasin comptent bien sûr dans la distance totale).
  - **des variables entières** (positives ou nulles) $\boldsymbol{charge_j}$ (j allant de 0 à n) représentent le nombre de vélos présents dans le semi-remorque. Le chargement initial du semi-remorque effectué au magasin est $charge_0$. Il reste a priori une inconnue du problème (on peut toutefois aisément le deviner s'il est possible de ré-équilibrer idéalement les stations. La variable $charge_j$ (pour $j \geq 1$) représente le nombre de vélos présent dans le semi-remorque après la $j^{\text{ème}}$ étape.
  - **des variables entières** (de signe quelconque) $\boldsymbol{depot_{ij}}$ qui représentent le nombre de vélos déposés à la station i (en cas de manque) ou retirés de la station i (en cas d'excédent) à l'étape j. On doit s'assurer que $depot_{ij}$ ne puisse pas prendre une valeur différente de 0 que dans le cas où la station i est la station effectivement visitée à l'étape j.

  Remarques importantes :

  - La distance totale parcourue par le semi-remorque, qui interviendra dans le modèle mathématique à écrire est intrinsèquement quadratique si l'on utilise les variables $\boldsymbol{x_{ij}}$. On pourra alors la linéariser afin d'obtenir un modèle linéaire en variables mixtes (certaines seront continues, d'autres 0-1 et d'autres entières).

- Le déséquilibre global est une somme de déséquilibres individuels, chaque déséquilibre individuel représentant une notion d'erreur. Il est naturel de représenter cette erreur au moyen d'une norme. La norme retenue ici étant la valeur absolue de l'écart à l'idéal (on aurait pu considérer une erreur quadratique). Il conviendra donc de pouvoir représenter de façon linéaire une valeur absolue.
- On cherchera à écrire un modèle unique permettant la prise en compte des deux objectifs.

3) Format des fichiers d'instances

Les instances à résoudre sont celles présentes dans le sharepoint du cours (fichier instances.zip).

L'exemple de base, illustré au paragraphe suivant est l'instance mini_6 :

```
name mini_6
K 6
stations
# id x y nbp capa ideal
1 82  9  8 11 8
2 69 79  3  9 7
3 71  9 10 10 7
4 64 93  2  6 4
5 65 13  5  7 5
whareouse 50 50
```

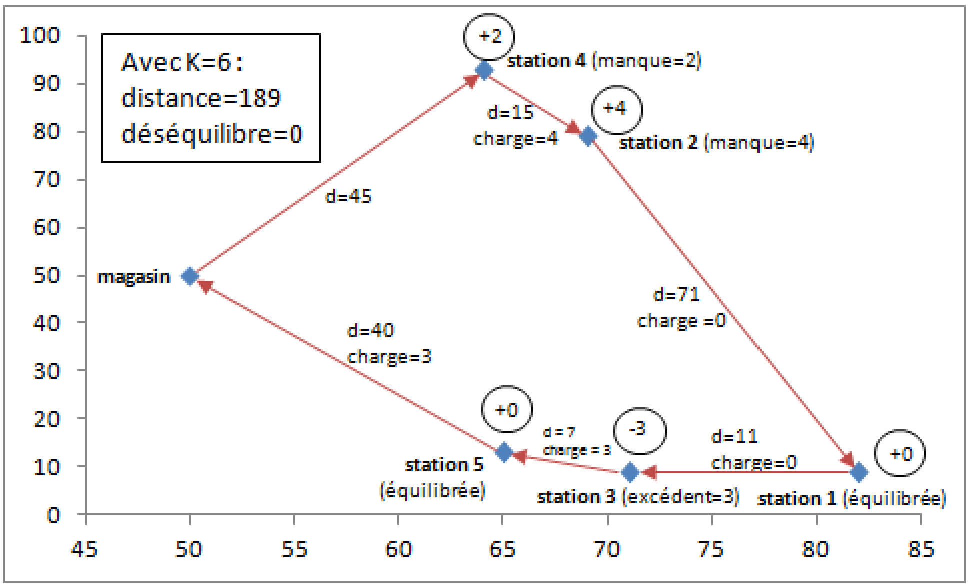
Fichier mini_6.dat

Tous les fichiers d'instances respectent les règles de syntaxe suivantes. Toute ligne commençant par '#' est un commentaire. La chaîne suivant le mot clé '`name`' est le nom de l'instance. Le nombre suivant le mot clé '`K`' est la capacité du semi-remorque. Les lignes suivant le mot clé '`stations`' et situées avant le mot clé '`warehouse`'correspondent aux caractéristiques des stations, avec pour chaque station une ligne indiquant son identifiant, son abscisse, son ordonnée, son 'nbp' (nombre de vélos présents avant la tournée), sa 'cap' (capacité de la station), et son 'ideal' (nombre idéal de vélos à atteindre après la tournée). Une dernière ligne contient le mot clé 'warehouse' suivi des coordonnées (x, y) du magasin.

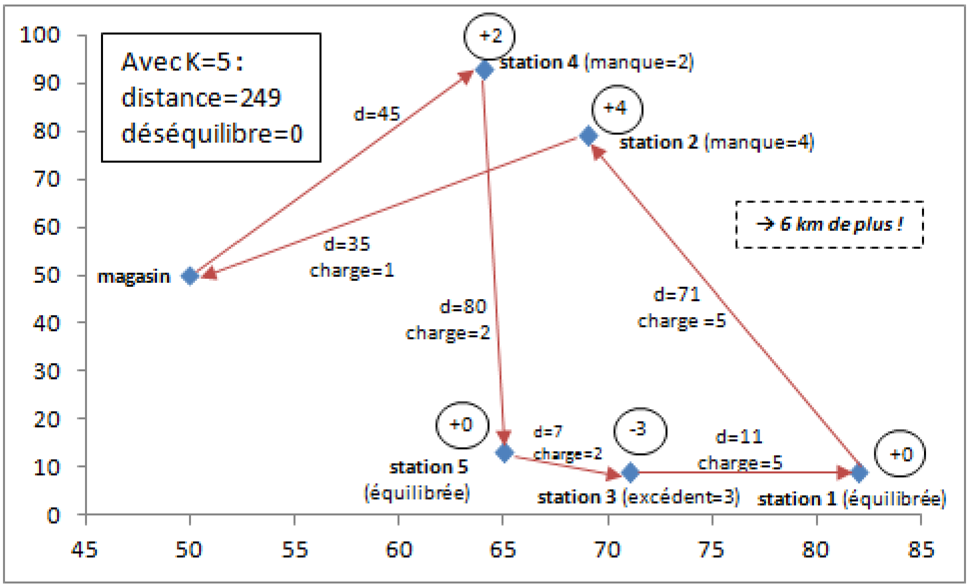
4) Trois exemples et leur solution

Pour illustrer le travail d'optimisation, nous allons considérer trois variantes de la même instance, l'instance mini_6 du paragraphe précédent, pour laquelle nous faisons varier la valeur du paramètre *K* (capacité du semi-remorque). Les instances mini_5 et mini_2 ne diffèrent donc de l'instance mini_6 que par la valeur de K (égale à 5 pour mini_5 et 2 pour mini_2). La résolution de ces trois instances fournit les solutions optimales suivantes.

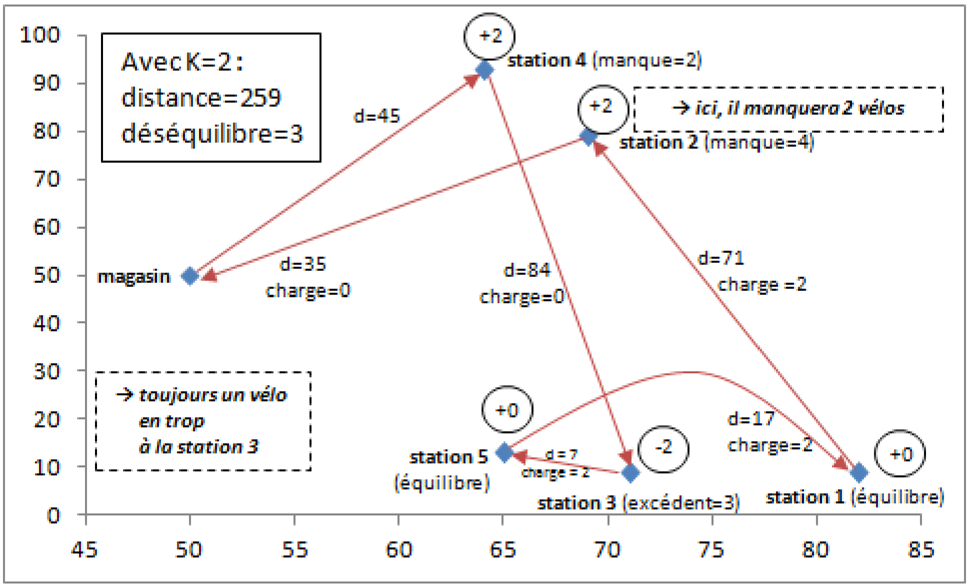
Légende :
- la mention entre parenthèses indique la situation initiale de la station
- x précise le nombre de vélos déposés ($x \geq 0$) ou retirés ($x \leq 0$) à la static
- le stock désigne le nombre de vélos présents dans le semi-remorque
- l'unité de distancce est l'hectomètre (1u=100m)



Solution optimale de l'instance mini_6



Solution optimale de l'instance mini_5



Solution optimale de l'instance mini_2

5) Travail demandé, format de sortie et échéancier

Le projet peut être réalisé en binôme. Le travail à réaliser est le suivant :

- **Modélisation** du problème. Le modèle mathématique permettant la résolution exacte du problème est à envoyer par email avant le 12/04/2025. Vous pouvez proposer une modélisation autre que celle qui est suggérée. Dans le cas d'un modèle quadratique, sa linéarisation est demandée. Un corrigé sera envoyé le 13/04, pour cette raison aucun retard ne pourra être accepté.

- Détermination d'une solution de meilleure qualité possible en utilisant une **méta-heuristique : le recuit simulé**, à programmer. La qualité des solutions fournies par votre méthode sera évaluée. Cette recherche heuristique d'une solution admissible de bonne qualité devrait s'avérer la seule voie possible de résolution pour les instances de grande taille.

- Détermination d'une solution par une méthode heuristique. Vous proposerez un algorithme de votre invention (autre que le recuit simulé et n'étant pas non plus une méta-heuristique connue de la littérature), le mettrez en œuvre dans le langage de votre choix.

- **Rédaction d'un rapport** indiquant les résultats obtenus, à rendre au plus tard le 31/05. Ce rapport devra en particulier :
    - expliquer la méta-heuristique développée, notamment en explicitant la notion de voisinage adoptée,
    - présenter les modèles adaptés du modèle initial en nombres entiers pour la détermination des bornes inférieures du déséquilibre et de la distance,
    - présenter les résultats numériques obtenus sur les instances à résoudre, par le recuit simulé et votre heuristique. Vous devrez comparer les valeurs numériques obtenues par les deux méthodes sur chacune des instances ainsi que le temps de résolution. Vous pouvez présenter les résultats en plusieurs tableaux pour une plus grande lisibilité.

Dates importantes :

- 03/04 : distribution du sujet
- 12/04, rendu de la modélisation du problème. Un modèle corrigé sera envoyé le 13/04
- 31/05 : rendu du rapport final