

Optimisation Discrète

PROJET – LIEUX DE FORMATION

THOMAS PERISSIER – JUSTINE GROLEAU – GWENDOLINE GOMEZ

Sommaire

I.	Etude du problème	2
II.	Choix d'une méthode.....	3
1.	Méthodes de trajectoire	3
2.	Méthodes de population.....	3
III.	Mise en place de l'algorithme	5
IV.	Etude des résultats.....	7
1.	Evolution de la durée de traitement selon la taille des données	7
2.	Evolution de la qualité de la solution selon le nombre d'itérations	8

I. Etude du problème

Une entreprise souhaite organiser des formations pour ses employés. Ceux-ci sont localisés dans des agences situées aux quatre coins de la France métropolitaine. Pour ce faire, elle a, à sa disposition, une salle dans de nombreuses villes en France afin de pouvoir réunir ses employés dans le cadre de ces formations.

Dans chaque agence, le nombre d'employés varie de un à dix. Chaque salle de formation peut accueillir 60 personnes au maximum. On considère que tous les employés d'une même agence se rendent dans une même salle. Les salles ont été choisies parmi les 1947 villes françaises de plus de 5 000 habitants.

Le coût de chaque salle comprend la rémunération des formateurs qui est de 2 000€, le coût de réservation de la salle comprenant le repas qui est de 1 000€ et le coût de transport de chaque employé des agences à la salle. Pour chaque personne, le trajet coûte 0.4€ par kilomètre parcouru. On considère qu'il n'y a pas de covoiturage possible. Il ne faut pas oublier que le trajet parcouru par chaque personne comprend l'aller ainsi que le retour.

Le problème posé est de réaliser un algorithme qui, pour un fichier contenant des agences données, va chercher à quelles salles affecter ces agences en réduisant les coûts au maximum.

On peut poser le coût d'une solution sous cette forme :

Soit x , les agences.

Soit n , le nombre d'agences.

Soit p_i le nombre de personne pour l'agence i .

Soit y , les salles de formation choisies.

Soit y_{xi} , la salle de formation associée à l'agence i .

Soit m , le nombre de salles de formation choisies.

Soit $d(x_i, y_j)$, la distance en kilomètres entre une agence i et une salle de formation j .

Coût d'une solution : $(2000 + 1000) * m + \sum_{i=1}^n 0.4 * p_i * d(x_i, y_{xi}) * 2$

II. Choix d'une méthode

Le problème posé est de réaliser une optimisation. Différentes approches ont été abordées en cours concernant la résolution de ce type de problème. On retrouve alors deux types de méthodes que l'on pourrait implémenter.

1. Méthodes de trajectoire

Les méthodes de trajectoire regroupent les méthodes qui partent d'une solution provisoire et qui essaient de l'améliorer par itération en modifiant une petite partie de la solution à chaque fois. Parmi ces méthodes, on trouve le recuit simulé et la recherche tabou.

- Le recuit simulé est issu de la thermodynamique. Le principe est alors d'accepter des solutions moins performantes quand on ne peut plus améliorer (minimal local) afin de chercher une solution un peu plus différente qui pourrait être meilleure. Nous avons besoin d'une température.
- La recherche tabou crée un voisinage qui correspond à des nouvelles solutions proches de la solution initiale. Elle garde ensuite la meilleure de solution. On peut utiliser une liste tabou dans laquelle on place les précédentes transformations afin de ne pas reboucler dessus.

2. Méthodes de population

Les méthodes de population regroupent les méthodes qui, à chaque itération, crée un ensemble de solutions en parallèle. Elles sont souvent comparées à un comportement naturel. Parmi ces méthodes, on trouve les algorithmes génétiques, les algorithmes de colonies de fourmis et l'optimisation par essaims particulaires.

- Les algorithmes génétiques sont empruntés à la théorie de l'évolution. Celle-ci sélectionne des mutations d'individus parmi ceux qui vivent et qui se perpétuent le mieux. Dans ces algorithmes, on sélectionne ceux qui répondent le mieux à la demande initiale. Dans notre cas, on sélectionnerait les couples agence/ville qui auraient les coûts les plus bas.
- Les algorithmes de colonies de fourmis sont issus du comportement de celles-ci lorsqu'elles trouvent de la nourriture : elles laissent une trace qui attire les autres afin qu'elles suivent le même chemin.
- L'optimisation par essaims particulaires s'inspire quant à elle du déplacement des groupes d'oiseaux. Elle revient à faire évoluer la solution en la faisant converger vers un minimum local grâce à des règles de déplacement simple.

Nous avons d'abord opté pour des méthodes de trajectoire car, les méthodes de population étant basées sur des comportements naturels, il nous semblait plus pertinent d'opter pour une méthode de trajectoire.

Nous avons ensuite choisi de sélectionner la méthode de recherche tabou car elle nous semblait plus cohérente avec notre problème : on part d'une solution initiale et petit à petit, agence par agence, on essaie d'améliorer la solution.

III. Mise en place de l'algorithme

Tout d'abord, il a fallu bien définir le terme de solution. Après une longue réflexion, nous nous sommes entendus sur le fait qu'une solution serait l'ensemble des agences où chaque agence serait associée à un lieu de formation. La première étape de l'algorithme est de créer une solution initiale. Celle-ci est créée en cherchant pour chaque agence la salle de formation disponible la plus proche. Cette méthode permet d'obtenir une solution initiale dont le coût est peu éloigné de celui de la solution finale.

Nous avons ensuite eu quelques difficultés à mettre en place l'algorithme de la méthode tabou. Notre première idée a été d'utiliser la liste tabou dans le cas où on ne trouvait pas de meilleure solution et où on dégradait la solution. L'idée était alors de parcourir les agences à chaque itération et de parcourir les salles parmi un échantillon au hasard afin de réduire le temps d'exécution.

On cherchait une salle qui réduirait les coûts en tenant compte des contraintes de place dans les salles et de distance. Si la recherche trouvait une meilleure salle, on l'assignait à l'agence. Sinon, on chercherait une salle un peu moins optimale que celle qui était assignée à l'agence et on plaçait cette dernière dans la liste tabou pour ne pas reboucler sur cette solution.

Après avoir implémenté cette méthode, nous nous sommes aperçus que malgré un grand nombre d'itérations, les meilleures solutions ne descendaient pas en dessous de 340 000€ pour la liste de 100 agences.

```
Cout actuel : 363695.894906195
Cout actuel : 356003.47012941766
Cout actuel : 371389.68734433345
Cout actuel : 364870.97562258446
Cout actuel : 363243.73141345975
Cout actuel : 395328.79400371714
Cout actuel : 373158.9288869674
Cout actuel : 379793.17749972193
Cout actuel : 376797.64460342424
Cout actuel : 369153.78432135144
Cout actuel : 380517.8522213912
Cout actuel : 381081.7653456859
Cout actuel : 364642.9317268816
BUILD STOPPED (total time: 15 minutes 18 seconds)
```

Nous avons alors décidé de repenser la mise en place de notre algorithme afin de trouver des résultats un peu plus optimisés. Nous nous sommes rendu compte que, de cette manière, nous changions trop d'élément de la solution : à chaque itération, chaque agence voyait la ville qui lui était assignée changer.

La mise en place de la solution initiale nous semblait cohérente donc nous l'avons conservé. A chaque itération, on parcourt les agences et les lieux de formation de l'échantillon afin d'en trouver un qui réduit le coût total. Lors de ce parcours, on gère l'ensemble des cas, c'est-à-dire, par exemple, l'augmentation du coût du transport entre une agence et un lieu de formation lorsque cette dernière n'augmente pas le coût total. A la fin du parcours des agences, on conserve cette solution qui va être la base d'une nouvelle itération améliorative.

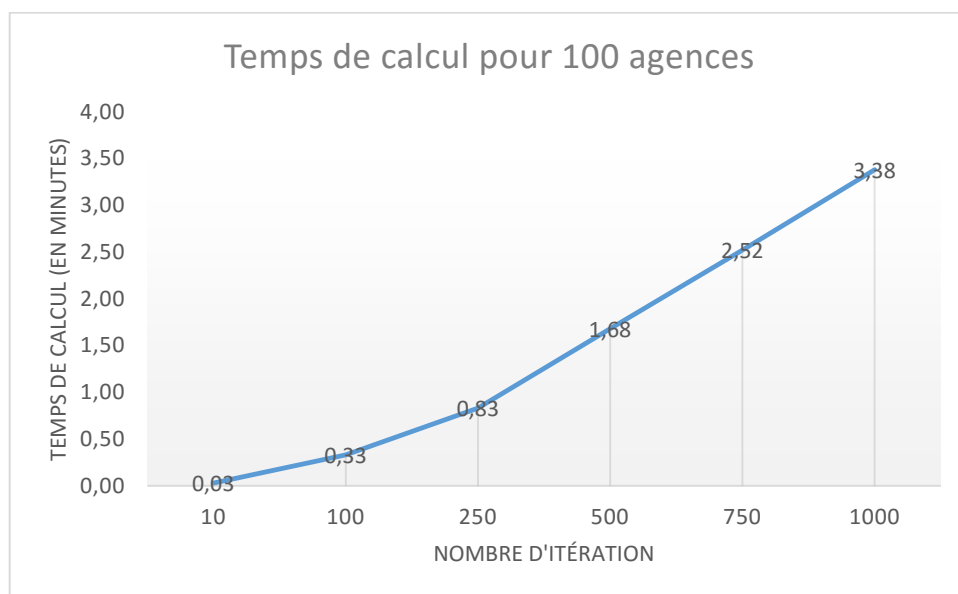
Nous nous sommes aperçu que, vu le nombre de possibilités, il y a peu de chance de reboucler sur les même solutions. Avec cette implémentation, nous arrivons à trouver des résultats entre 110 000 et 120 000€ pour la liste de 100 agences et avec 1 000 itérations. Nous avons donc considérablement amélioré notre méthode.

Après 1 000 itérations en 4 minutes de traitement

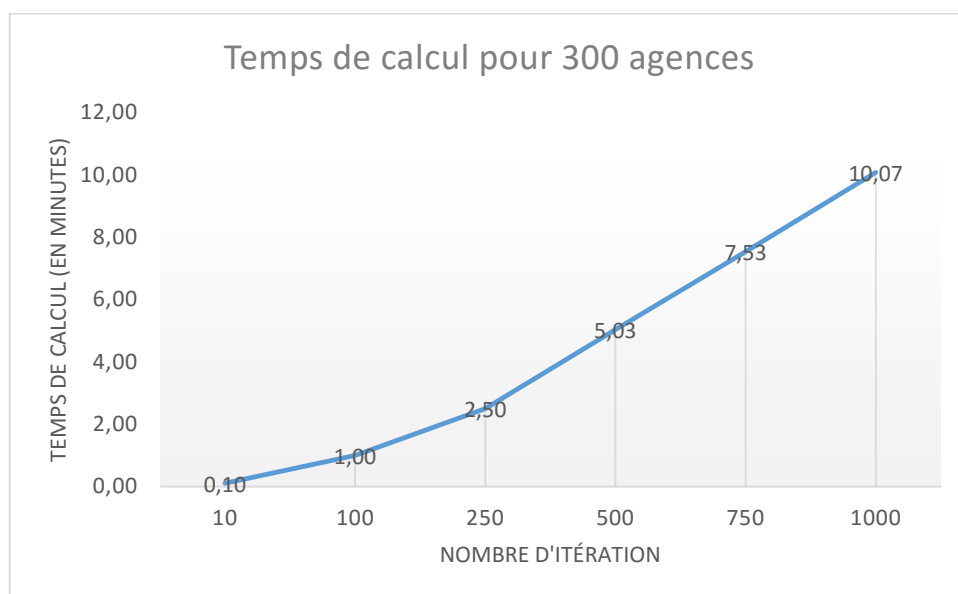
```
Nombre LF de la solution initiale:81
Cout LF : 243000.0
Cout Transport : 7696.141222028669
Cout total: 250696.14122202867
-----
Nombre LF de la solution finale:23
Cout LF : 69000.0
Cout Transport : 38083.03299524495
Cout total: 107083.03299524495
```

IV. Etude des résultats

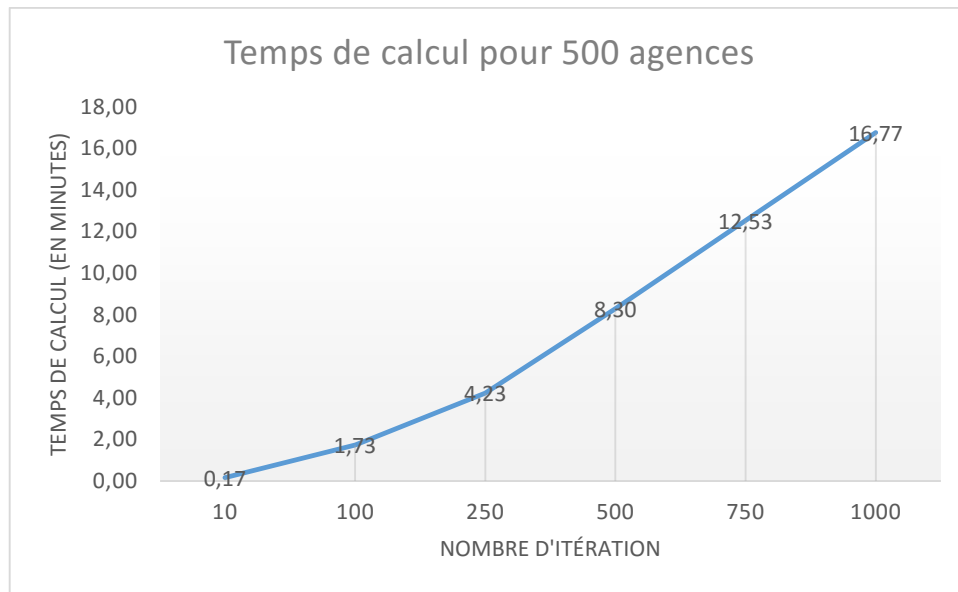
1. Evolution de la durée de traitement selon la taille des données



Graphe des temps de calcul pour 100 agences

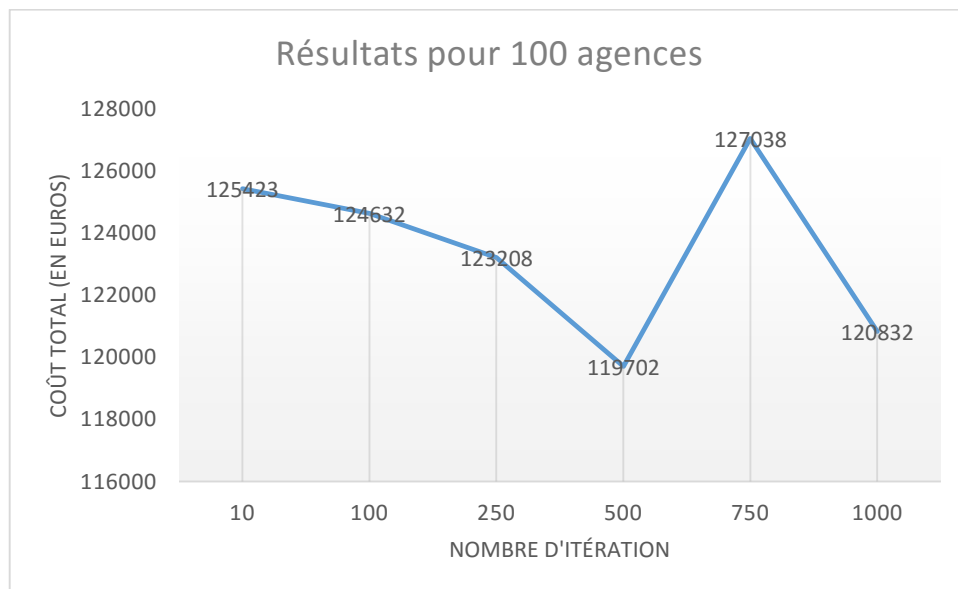


Graphe des temps de calcul pour 300 agences

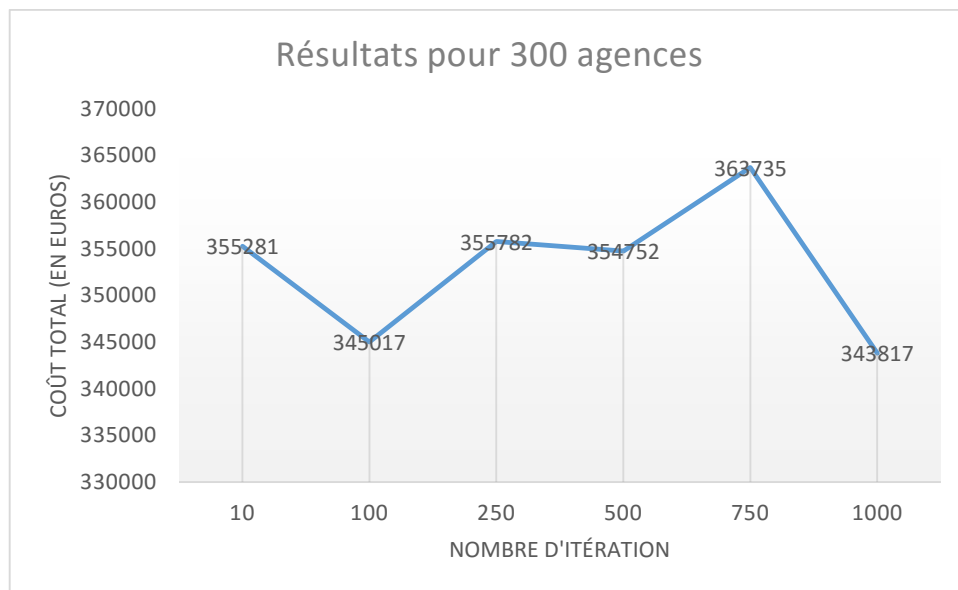


Graphe des temps de calcul pour 500 agences

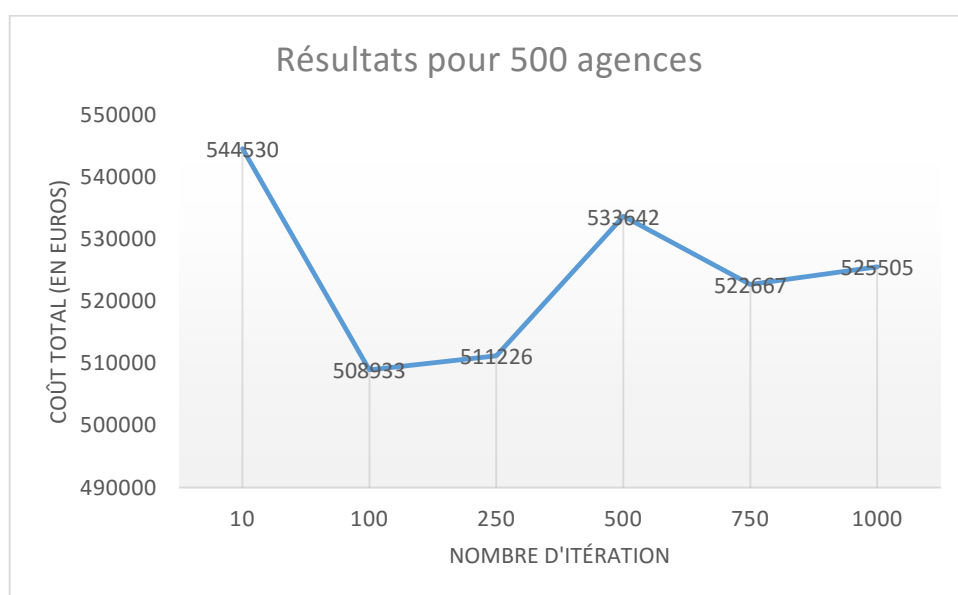
2. Evolution de la qualité de la solution selon le nombre d'itérations



Graphe des résultats pour 100 agences



Grphe des résultats pour 300 agences



Grphe des résultats pour 500 agences