

INF431

Graphes eulériens et problème du facteur chinois Sujet proposé par Gilles Schaeffer

Version: 4:2459M

Un postier se prépare à parcourir toutes les rues d'un village pour distribuer le courrier ; en fonction du plan des rues il risque de devoir passer plusieurs fois dans la même rue, et il se pose la question de minimiser la longueur de son trajet.

Nous modélisons ce problème comme un problème d'*optimisation sur un graphe* : il s'agit de trouver dans un graphe dont les arêtes sont pondérées, un cycle de poids minimal parmi les cycles qui visitent toutes les arêtes. Lorsque chaque sommet est incident à un nombre pair d'arêtes, nous verrons qu'une approche gloutonne peut donner le résultat en temps linéaire. D'un autre côté dès qu'il y a des sommets incidents à un nombre impair d'arêtes, l'approche gloutonne échoue et il n'est pas évident que ce problème soit plus simple que, par exemple, le problème du voyageur de commerce, notoirement difficile (NP-complet).

Pourtant nous allons voir que le problème peut être résolu en temps polynomial par une analyse combinatoire assez fine : c'est à la fois le charme ou le défaut (suivant les goûts) de ces questions d'optimisation sur les graphes que de souvent nécessiter une étude ad-hoc astucieuse pour savoir si le problème est facile (polynomial) ou difficile (NP-complet)... Pour conclure cette introduction, remarquons que la situation est vraiment compliquée : le même problème reste polynomial sur un graphe orienté, mais devient difficile (NP-complet) sur un graphe mélangeant arêtes orientées et non-orientées. D'où l'importance dans ce domaine du paradigme polytechnicien : savoir reconnaître un problème connu et s'y ramener...

1 Cycles et chemins eulériens

L'étude des graphes eulériens est souvent associée à une généralisation de la notion de graphe, où deux sommets peuvent être reliés par plus d'une arête. Nous utiliserons la définition suivante :

Définition 1 (multigraphe) Un *multigraphe non orienté sans boucle* $\mathcal{G} = (S, A)$ est composé d'un ensemble de sommets S et d'un ensemble d'arêtes A . Chaque arête $a \in A$ est de la forme $(i, \{s, s'\})$, où i est un identifiant unique pris parmi un ensemble d'identifiants I , et où s et s' sont deux sommets distincts. \diamond

(Dans la suite, on dira parfois « graphe » au lieu de « multigraphe ».) La notion de chemin est adaptée à ce cadre :

Définition 2 (chemin) Un *chemin* P dans \mathcal{G} est une liste alternée $(s_0, i_1, s_1, i_2, \dots, i_k, s_k)$ de sommets et d'identifiants d'arêtes, commençant et terminant par un sommet ($k \geq 0$), telle que, pour tous (s, i, s') consécutifs, $(i, \{s, s'\})$ est une arête de \mathcal{G} . On note $|P| = k$ la *longueur*, c'est-à-dire le nombre d'arêtes, d'un tel chemin P . \diamond

Définition 3 (concaténation) Si P et P' sont deux chemins tels que l'extrémité droite de P coïncide avec l'extrémité gauche de P' , alors $P \# P'$ dénote le chemin obtenu par concaténation de P et P' . \diamond

Définition 4 (cycle) Un *cycle* est un chemin dont les sommets extrémaux sont identiques. \diamond

Définition 5 (chemin eulérien, cycle eulérien) Un chemin (ou un cycle) est dit *eulérien* s'il emprunte une et une seule fois chaque arête. Un multigraphe possédant un cycle eulérien est lui-même dit eulérien. \diamond

On notera qu'un chemin eulérien peut passer plusieurs fois par un même sommet.

Définition 6 (chemin simple, cycle simple) Un chemin (ou, par extension, un cycle) est dit *simple* s'il n'emprunte pas deux fois la même arête (même en des sens opposés) et s'il contient au moins une arête. \diamond

L'énoncé suivant caractérise de façon particulièrement simple l'existence d'un cycle eulérien. Dans la suite, nous démontrerons cet énoncé de façon constructive.

Théorème 1 (Euler, 1736) Un multigraphe connexe possède un cycle eulérien si et seulement si chaque sommet est de degré pair. \diamond

Construction efficace d'un cycle eulérien

Dans cet exercice, on se donne un multigraphe $\mathcal{G} = (S, A)$. On pose $n = |S|$ et $m = |A|$.

Question 1 On suppose que tous les sommets de \mathcal{G} sont de degré pair et qu'il existe un chemin $P = (s_0, i_1, \dots, i_k, s_k)$ passant au plus une fois par chaque arête. Montrer que soit P est un cycle, soit il peut être prolongé en un cycle qui emprunte au plus une fois chaque arête. (Toutefois, ce cycle n'emprunte pas nécessairement toutes les arêtes du multigraphe, et n'est donc pas en général un cycle eulérien.) \diamond

Question 2 On suppose que tous les sommets de \mathcal{G} sont de degré pair. Écrire le pseudo-code d'une fonction `FINDCYCLE(s)` qui construit et renvoie un cycle, issu du sommet s , passant au plus une fois par chaque arête. Cette fonction retirera du multigraphe les arêtes empruntées par ce cycle. Elle fera en sorte de produire un cycle suffisamment long pour que, une fois ces arêtes retirées, le sommet s soit isolé (un sommet est isolé s'il n'est incident à aucune arête). \diamond

Il reste à déterminer comment, à l'aide de `FINDCYCLE`, on construit un cycle eulérien.

Question 3 On suppose que \mathcal{G} est connexe et que tout sommet de \mathcal{G} est de degré pair. Écrire le pseudo-code d'une fonction `EULER()` qui construit et renvoie un cycle eulérien de \mathcal{G} . Quelle en est la complexité ? \diamond

Nous avons ainsi démontré constructivement le théorème d'Euler.

Question 4 Proposer une structure de données pour représenter les graphes qui permette d'implémenter l'algorithme `FINDCYCLE` (et donc `EULER`) en temps linéaire. \diamond

2 Le problème du facteur chinois

On étudie maintenant un problème d'optimisation appelé *problème du facteur chinois* ou *Chinese postman problem*. Il a été posé par le mathématicien chinois Mei Ko Kwan en 1962. Ce problème dérive de celui dit des ponts de Königsberg, résolu par Euler en 1736.

Présentation du problème d'optimisation

Soit $\mathcal{G} = (S, A)$ un multigraphe non orienté connexe, et $\text{coût}(\cdot) : A \rightarrow \mathbb{N}$ une valuation des arêtes de \mathcal{G} . On notera que deux arêtes reliant la même paire de sommets peuvent avoir des valuations différentes. On suppose par la suite que \mathcal{G} ne contient pas de cycle simple de valuation nulle.

Le *problème du facteur chinois* consiste à construire un *cycle de valuation minimale* passant au moins une fois par chaque arête. La figure 1 propose un exemple de problème. La métaphore du facteur considère les arêtes comme les rues d'un quartier et la valuation comme la distance entre deux carrefours : le facteur, partant du bureau de poste, doit parcourir toutes les rues et revenir au bureau de poste tout en minimisant la distance totale.

Bien entendu, si \mathcal{G} est eulérien, alors tout cycle eulérien est solution optimale à ce problème, et ce quelle que soit la valuation $\text{coût}(\cdot)$. Dans la suite on appelle *tournée* un cycle passant au moins une fois par chaque arête.

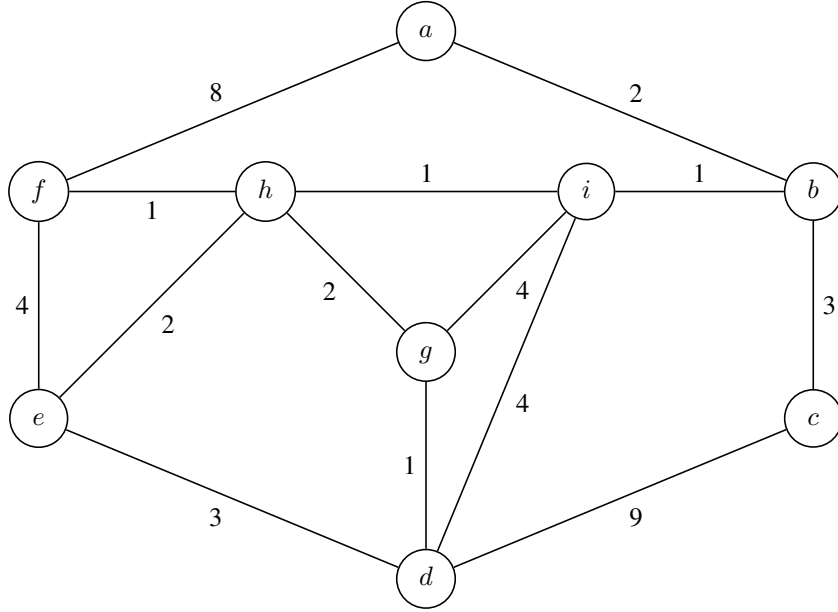


FIG. 1 – Une instance du problème du facteur chinois

Si l'on parcourt un cycle eulérien en dessinant les arêtes au fur et à mesure du parcours, on reconstitue le graphe eulérien sous-jacent. Plus généralement, si l'on parcourt une *tournée* C d'un graphe \mathcal{G} , qui emprunte éventuellement plusieurs fois certaines arêtes, en dessinant *de nouvelles arêtes* au fur et à mesure du parcours, ces nouvelles arêtes forment un *multigraphe eulérien* \mathcal{G}' , dans lequel les arêtes empruntées plusieurs fois par la tournée ont été *dupliquées*. (C'est ici que nous exploitons réellement la notion de multigraphe.) Nous allons utiliser cette construction pour ramener l'étude des tournées à celle de cycles eulériens. Le lemme suivant formalise la construction esquissée ci-dessous et pourra être utile pour la rédaction de preuves formelles des résultats suivants.

Lemme 1 Le graphe $\mathcal{G} = (S, A)$ admet une tournée C de valuation c si et seulement si il existe un cycle eulérien C' de valuation c dans un graphe $\mathcal{G}' = (S, A')$ muni d'une surjection $\phi : A' \rightarrow A$ qui préserve les extrémités des arêtes et leurs valuations. \diamond

Preuve : Soit $C = (s_0, i_1, \dots, i_k, s_k)$ une tournée de \mathcal{G} . Notons que, en général, les identifiants i_j ne sont pas deux à deux distincts, puisque C n'est pas un cycle eulérien. Nous allons choisir des identifiants uniques, par exemple les entiers de 1 à k , pour obtenir un cycle eulérien. Soit donc C' le chemin $(s_0, 1, \dots, k, s_k)$. Le chemin C' est un cycle eulérien vis-à-vis du multigraphe $\mathcal{G}' = (S, A')$, où A' est l'ensemble des arêtes parcourues par C' , à savoir $\{(j, \{s_{j-1}, s_j\}) \mid 1 \leq j \leq k\}$. La fonction $\phi : A' \rightarrow A$ qui à l'arête d'identifiant j associe l'arête d'identifiant i_j est surjective : en effet, puisque C est une tournée, l'ensemble $\{i_1, \dots, i_k\}$ contient (les identifiants de) *toutes* les arêtes de \mathcal{G} .

Réciproquement, soit C' un cycle eulérien d'un graphe \mathcal{G}' muni d'une telle surjection ϕ . Soit $C = \phi(C')$. La conservation des extrémités des arêtes par ϕ assure que C est un cycle de \mathcal{G} . La conservation des valuations assure que C' et C ont même valuation. Enfin, la surjectivité de ϕ garantit que toutes les arêtes de \mathcal{G} sont empruntées au moins une fois par C , donc que C est une tournée. \square

Sur la structure de l'ensemble des arêtes réutilisées dans une tournée

Question 5 Montrer que, parmi les tournées de valuation minimale sur un graphe \mathcal{G} , il en existe une qui emprunte au plus deux fois chaque arête. \diamond

Il est donc licite de rechercher la solution du problème du facteur chinois parmi les tournées empruntant chaque arête de \mathcal{G} au plus deux fois. Pour les trois questions suivantes, on fixe une telle tournée C ,

de valuation minimale. On note D l'ensemble des arêtes de \mathcal{G} empruntées deux fois dans C , et on étudie le graphe $\mathcal{D} = (S, D)$ formé par ces arêtes. Notons que ces définitions impliquent $|C| = |D| + |A|$.

Question 6 Montrer, en utilisant le graphe \mathcal{G}' , que le graphe \mathcal{D} ne contient pas de cycle simple. \diamond

Question 7 Montrer que les chemins simples du graphe \mathcal{D} sont des plus courts chemins du graphe \mathcal{G} . (On pourra se contenter de décrire comment adapter la preuve de la question précédente.) \diamond

Question 8 Montrer que, pour tout sommet s , la parité du degré de s dans \mathcal{G} est égale à la parité du degré de s dans \mathcal{D} . En déduire qu'on peut partitionner l'ensemble d'arêtes D en un ensemble de chemins simples dont les extrémités sont toutes distinctes et sont formées de l'ensemble des sommets de degré impair de \mathcal{G} . \diamond

Réduction à un problème de couplage parfait de poids minimum

Dans le but de résoudre le problème du facteur chinois, on cherche donc à apparier (c'est-à-dire à relier deux à deux) tous les sommets de \mathcal{G} de degré impair, et ce au moyen de plus courts chemins.

On construit un graphe auxiliaire non orienté \mathcal{G}_{aux} , dont les sommets sont les sommets de degré impair de \mathcal{G} . Le graphe \mathcal{G}_{aux} est *complet* : toute paire de sommets distincts forme une arête. La valuation associée à une arête $\{s, s'\}$ de \mathcal{G}_{aux} est la valuation d'un plus court chemin entre s et s' dans \mathcal{G} .

L'algorithme de Floyd-Warshall [1] permet de construire l'ensemble des plus courts chemins de \mathcal{G} en temps $\mathcal{O}(n^3)$. Si le nombre k de sommets de degré impair est faible devant n , ou bien si le nombre m d'arêtes du graphe est faible devant n^2 , on peut préférer appliquer l'algorithme de Dijkstra pour calculer les distances à partir de chacun des sommets de degré impair, pour une complexité totale de $\mathcal{O}(k \cdot (n + m) \log n)$.

Question 9 Construire le graphe auxiliaire, dans le cas où \mathcal{G} est le graphe de la figure 1. \diamond

Définition 7 (couplage parfait) Un *couplage* d'un graphe \mathcal{G} est un sous-graphe de \mathcal{G} (comportant donc les mêmes sommets, et en général moins d'arêtes, que \mathcal{G}) dont chaque sommet appartient à au plus une arête. Un couplage est dit *parfait* lorsque chaque sommet appartient à une arête. \diamond

Question 10 Montrer que \mathcal{G}_{aux} admet un couplage parfait. \diamond

Question 11 Montrer que le problème du facteur chinois se réduit à la recherche d'un couplage parfait de valuation minimale pour le graphe \mathcal{G}_{aux} . \diamond

Question 12 Trouvez un couplage parfait de valuation minimale pour le graphe auxiliaire construit à la question 9, et déduisez-en une solution optimale au problème du facteur chinois pour le graphe valué de la figure 1. \diamond

Dans le cas eulerien la complexité de la construction d'une tournée optimale est linéaire en m . Pour k fixé (petit) la complexité résultante est à peine moins bonne : nous avons perdu un facteur $\log n$ lors du calcul des plus courts chemins. Pour des valeurs de k arbitraires il faut par contre tenir compte du temps de calcul du couplage parfait de valuation minimale : Le meilleur algorithme connu pour déterminer un couplage parfait de valuation minimale est dû à Gabow [2]. Cet algorithme dépasse largement le cadre de ce cours ; sa complexité asymptotique pour un graphe à k sommets et ℓ arêtes est $\mathcal{O}(k(\ell + k \log k))$. Edmonds a donné le premier algorithme polynomial pour résoudre ce problème classique : sa complexité était de $\mathcal{O}(\ell k^2)$. Une bonne introduction aux problèmes de couplage est disponible dans l'ouvrage de référence de Gondran et Minoux [3]. Dans notre cas $\ell = \Theta(k^2)$ – le graphe auxiliaire est complet – de sorte que l'algorithme de Gabow donne une complexité $\mathcal{O}(k^3)$.

Références

[1] Thomas H. Cormen, Charles E. Leiserson, Ronald L. Rivest et Clifford Stein. *Introduction à l'algorithmique : Cours et exercices*. Sciences Sup. Dunod, 2002.

- [2] Harold N. Gabow. [Data structures for weighted matching and nearest common ancestors with linking](#). In *ACM-SIAM Symposium on Discrete Algorithms (SODA)*, pages 434–443, 1990.
- [3] M. Gondran et M. Minoux. *Graphes et algorithmes*. Eyrolles, 1995.