Université de Montréal

Tarification logit dans un réseau

par
François Gilbert

Département d'informatique et de recherche opérationnelle
Faculté des arts et des sciences

Thèse présentée à la Faculté des arts et des sciences
en vue de l'obtention du grade de Philosophiæ Doctor (Ph.D.)
en informatique

Décembre, 2011

Université de Montréal
Faculté des arts et des sciences

Cette thèse intitulée:

Tarification logit dans un réseau

présentée par:

François Gilbert

a été évaluée par un jury composé des personnes suivantes:

Fabian Bastin,	président-rapporteur
Patrice Marcotte,	directeur de recherche
Gilles Savard,	codirecteur
Jean-Yves Potvin,	membre du jury
David Watling,	examinateur externe
Lars Ehlers,	représentant du doyen de la FAS

Thèse acceptée le: 4 avril 2012

RÉSUMÉ

Le problème de tarification qui nous intéresse ici consiste à maximiser le revenu généré par les usagers d'un réseau de transport. Pour se rendre à leurs destinations, les usagers font un choix de route et utilisent des arcs sur lesquels nous imposons des tarifs. Chaque route est caractérisée (aux yeux de l'usager) par sa « désutilité », une mesure de longueur généralisée tenant compte à la fois des tarifs et des autres coûts associés à son utilisation. Ce problème a surtout été abordé sous une modélisation déterministe de la demande selon laquelle seules des routes de désutilité minimale se voient attribuer une mesure positive de flot. Le modèle déterministe se prête bien à une résolution globale, mais pèche par manque de réalisme. Nous considérons ici une extension probabiliste de ce modèle, selon laquelle les usagers d'un réseau sont alloués aux routes d'après un modèle de choix discret logit. Bien que le problème de tarification qui en résulte est non linéaire et non convexe, il conserve néanmoins une forte composante combinatoire que nous exploitons à des fins algorithmiques.

Notre contribution se répartit en trois articles. Dans le premier, nous abordons le problème d'un point de vue théorique pour le cas avec une paire origine-destination. Nous développons une analyse de premier ordre qui exploite les propriétés analytiques de l'affectation logit et démontrons la validité de règles de simplification de la topologie du réseau qui permettent de réduire la dimension du problème sans en modifier la solution. Nous établissons ensuite l'unimodalité du problème pour une vaste gamme de topologies et nous généralisons certains de nos résultats au problème de la tarification d'une ligne de produits.

Dans le deuxième article, nous abordons le problème d'un point de vue numérique pour le cas avec plusieurs paires origine-destination. Nous développons des algorithmes qui exploitent l'information locale et la parenté des formulations probabilistes et déterministes. Un des résultats de notre analyse est l'obtention de bornes sur l'erreur commise par les modèles combinatoires dans l'approximation du revenu logit. Nos essais numériques montrent qu'une approximation combinatoire rudimentaire permet souvent d'identifier des solutions quasi-optimales.

Dans le troisième article, nous considérons l'extension du problème à une demande hétérogène. L'affectation de la demande y est donnée par un modèle de choix discret logit mixte où la sensibilité au prix d'un usager est aléatoire. Sous cette modélisation, l'expression du revenu n'est pas analytique et ne peut être évaluée de façon exacte. Cependant, nous démontrons que l'utilisation d'approximations non linéaires et combinatoires permet d'identifier des solutions quasi-optimales. Finalement, nous en profitons pour illustrer la richesse du modèle, par le biais d'une interprétation économique, et examinons plus particulièrement la contribution au revenu des différents groupes d'usagers.

Mots clés: conception de réseau, modèles de choix discrets, programmation biniveau, optimisation combinatoire, optimisation non linéaire.

ABSTRACT

The network pricing problem consists in finding tolls to set on a subset of a network's arcs, so to maximize a revenue expression. A fixed demand of commuters, going from their origins to their destinations, is assumed. Each commuter chooses a path of minimal "disutility", a measure of discomfort associated with the use of a path and which takes into account fixed costs and tolls. A deterministic modelling of commuter behaviour is mostly found in the literature, according to which positive flow is only assigned to "shortest" paths. Even though the determinist pricing model is amenable to global optimization by the use of enumeration techniques, it has often been criticized for its lack of realism. In this thesis, we consider a probabilistic extension of this model involving a logit dicrete choice model. This more realistic model is non-linear and non-concave, but still possesses strong combinatorial features.

Our analysis spans three separate articles. In the first we tackle the problem from a theoretical perspective for the case of a single origin-destination pair and develop a first order analysis that exploits the logit assignment analytical properties. We show the validity of simplification rules to the network topology which yield a reduction in the problem dimensionality. This enables us to establish the problem's unimodality for a wide class of topologies. We also establish a parallel with the product-line pricing problem, for which we generalize some of our results.

In our second article, we address the problem from a numerical point of view for the case where multiple origin-destination pairs are present. We work out algorithms that exploit both local information and the pricing problem specific combinatorial features. We provide theoretical results which put in perspective the deterministic and probabilistic models, as well as numerical evidence according to which a very simple combinatorial approximation can lead to the best solutions. Also, our experiments clearly indicate that under any reasonable setting, the logit pricing problem is much smoother, and admits less optima then its deterministic counterpart.

The third article is concerned with an extension to an heterogeneous demand resulting from a mixed-logit discrete choice model. Commuter price sensitivity is assumed random and the corresponding revenue expression admits no closed form expression. We devise nonlinear and combinatorial approximation schemes for its evaluation and optimization, which allow us to obtain quasi-optimal solutions. Numerical experiments here indicate that the most realistic model yields the best solution, independently of how well the model can actually be solved. We finally illustrate how the output of the model can be used for economic purposes by evaluating the contributions to the revenue of various commuter groups.

Keywords: network design, discrete choice models, mixed-logit, bilevel programming, combinatorial optimization, nonlinear optimization.

TABLE DES MATIÈRES

LISTE DES TABLEAUX

LISTE DES FIGURES

CHAPITRE 1

INTRODUCTION

Le problème qui consiste à maximiser les revenus générés par l'imposition de tarifs sur un sous-ensemble des arcs d'un réseau a fait l'objet de plusieurs travaux, principalement dans un contexte déterministe où tous les usagers associés à un même couple origine-destination sont affectés à un unique plus court chemin, résultant en une affectation « tout-ou-rien ». Dans cette thèse, nous considérons des modèles plus réalistes qui tiennent compte de la stochasticité et de l'hétérogénéité de la demande. Ces modèles, de par leur double nature combinatoire et continue, posent des défis algorithmiques considérables auxquels nous apporterons des solutions efficaces. En particulier, nous développerons des algorithmes permettant de déterminer des solutions quasi-optimales pour des instances de tailles non triviales.

1.1 Le problème de tarification déterministe

Dans le modèle de base, introduit par Labbé et al. [51], deux critères sont associés à chaque arc a du réseau : un coût fixe c_a et un tarif t_a. Étant donné un ensemble d'arcs tarifés $\mathscr{A}_{\text{toll}}$, on a bien sûr $t_a = 0$ si $a \notin \mathscr{A}_{\text{toll}}$. Pour chaque couple origine-destination, la demande (fixe, donc indépendante des tarifs) est affectée sur un chemin de plus faible désutilité (ou « coût »), où la désutilité d'un chemin r s'exprime comme la somme des désutilités des arcs (somme de coûts fixes et variables) qui composent ce chemin, c'est-à-dire :

$$u_r(t) = \sum_{a \in r} c_a + \alpha\, t_a. \tag{1.1}$$

Dans le cas où le coût fixe c_a représente un délai, le paramètre α représente la valeur monétaire d'une unité de délai (« valeur du temps »). Dans nos deux premières publications, ce paramètre de sensibilité est supposé constant, c'est-à-dire que la population est homogène. Dans notre troisième publication, il est traité comme une variable aléatoire, et permet donc de modéliser la variation de la sensibilité au prix dans la population.

Le problème de tarification déterministe se formule naturellement comme un programme mathématique bi-niveau [14, 25, 27, 51]. Les contraintes du programme imposent qu'un sous-ensemble de variables forme une solution optimale d'un programme mathématique auxiliaire : le problème de niveau inférieur. Dans notre cas, le niveau inférieur permet de modéliser la réaction des usagers à la politique de tarification t, fixée au premier niveau par le « meneur ». La solution optimale x du problème de niveau inférieur est une affectation des flots sur les chemins $\mathscr{R}$ du réseau. On a

Programme 1.1.

$$\begin{aligned}
\max_{t,x} \quad & \sum_{r\in\mathscr{R}} (x_r \sum_{a\in r} t_a) \\
\textit{sujet à} \quad & t_a = 0 \textit{ si } a \notin \mathscr{A}_{\text{toll}} \\
& x \in \arg\min_{y} \left\{ \sum_{r\in\mathscr{R}} y_r\, u_r(t) : \sum_{r\in\mathscr{R}} y_r = 1, y \geq 0 \right\}.
\end{aligned}$$

La résolution exacte de ce type de programme fait en général intervenir une reformulation linéaire en nombres entiers, définie sur un seul niveau, qui est obtenue en substituant au niveau inférieur ses conditions d'optimalité. Les solutions optimales du problème de niveau inférieur correspondent à des points extrémaux dans un polyèdre défini dans l'espace primal-dual. Un algorithme de type séparation et évaluation progressive tire alors profit de cette structure combinatoire du problème. Des instances de tailles limitées peuvent être traitées par cette approche ; des heuristiques que l'on adapte spécialement à la structure du problème permettent d'aborder les instances de grande taille. La plupart de ces heuristiques font appel au problème d'optimisation inverse, consistant à déterminer quels sont les tarifs optimaux associés à un ensemble de flots fixés au niveau inférieur d'optimisation, et qui se ramène dans notre cas à un programme linéaire.

Par exemple, Dussault et al. proposent de lisser les contraintes de complémentarité associées aux conditions d'optimalité du niveau inférieur dans [36], pour un problème de tarification dont la structure de base n'est pas nécessairement réticulaire. En diminuant progressivement le degré de lissage, on obtient une trajectoire dans l'espace des tarifs dont les programmes deviennent de moins en moins bien conditionnés. Chaque nouveau vecteur de tarifs est la solution du problème d'optimisation inverse, lui-même défini sur la base des flots optimaux de l'itération précédente. Le bon comportement de la méthode est attesté par des essais numériques qui produisent fréquemment des solutions optimales ou quasi-optimales.

Dans le contexte plus général de la programmation bi-niveau non linéaire avec contraintes linéaires, Marcotte et al. [57] présentent un algorithme de région de confiance. Plutôt que d'utiliser un modèle quadratique traditionnel, l'algorithme de Marcotte et al. se base sur un modèle combinatoire, obtenu par par le biais d'approximations d'ordre un ou deux des objectifs des niveaux inférieur et supérieur de la formulation exacte (sur deux niveaux). L'application de cette technique au problème de tarification déterministe par Colson et al. [26] a permis de constater que la suite des itérés produits par l'algorithme de région de confiance bi-niveau est beaucoup moins enclin à converger trop rapidement vers un optimum local (qui n'est pas un optimum global) que celle produite par une méthode de montée traditionnelle. Cette approche sera adaptée au modèle de tarification logit dans notre seconde publication sous une forme impliquant des modèles combinatoires plus élaborés.

À partir d'une reformulation primale-duale bilinéaire obtenue en remplaçant le problème de niveau inférieur par ses conditions d'optimalité, Brotcorne et al. [15] pénalisent quadratiquement un sous-ensemble de contraintes dans l'objectif. Le programme mathématique ainsi obtenu se prête bien à l'optimisation successive par rapport à deux groupes de variables choisis de façon appropriée. Couplé à l'optimisation inverse, ce schéma gaus-seidelien permet d'obtenir de bonnes solutions pour des problèmes de tarification de grande taille. Nous utiliserons une approche analogue pour la résolution de certains modèles dans notre deuxième article, et notamment dans le cadre d'un algorithme par région de confiance.

Bouhtou et al. [12] proposent une procédure de simplification amenant une réduction de la dimension du problème dans laquelle intervient la notion de « chemin dominé ». Un chemin est dominé s'il ne peut pas être le plus court sous toute tarification admissible ; il peut donc être ignoré. Nous généralisons ce type d'analyse au cas d'une affectation logit dans notre premier article et nous y démontrons la validité de règles de simplification de la topologie du réseau laissant la valeur des optima inchangée.

1.2 Le modèle de choix discret logit

Malheureusement, l'affectation des usagers aux plus courts chemins uniquement pèche par manque de réalisme. En effet, on peut démontrer qu'il y aura toujours au moins deux chemins possédant une désutilité identique à l'optimum et qu'il n'est pas raisonnable de supposer qu'un seul chemin soit utilisé. Cette remarque vaut également pour des chemins dont les désutilités sont proches. Pour pallier à ce problème, nous considérons un modèle d'affectation de type « choix discret », le plus simple étant le modèle logit, qui se justifie d'au moins trois manières : (i) sur une base axiomatique [53] , (ii) par la théorie de l'information [31, 37] ou (iii) par la théorie de l'utilité aléatoire [58]. L'approche axiomatique impose aux proportions formées par les différentes probabilités de choix, lorsque prises deux à deux, de demeurer inchangées suite à l'ajout de nouvelles « alternatives ». Ceci correspond à la propriété d'indépendance des alternatives non pertinentes IIA (*Independance of Irrelevant Alternatives* , en anglais). La distribution logit est la seule qui satisfasse cet axiome. L'approche par la théorie de l'information situe l'affectation logit dans un cadre statistique comme solution d'un problème d'entropie. En faisant appel au principe de raison insuffisante, la distribution logit se présente comme la distribution la plus probable, étant donné que la désutilité espérée du parcours choisi n'excède pas une certaine borne. On se réfère le plus souvent à l'approche par la théorie de l'utilité aléatoire qui fait intervenir la notion d'un « décideur » cherchant à maximiser son utilité au moment de faire son choix. Les utilités des alternatives sont modélisées par des variables aléatoires indépendantes de loi Gumbel dont les espérances sont exprimées par des fonctions linéaires d'attributs (dans notre cas : les coûts fixes et les tarifs imposés sur une route) pondérés de façon appropriée au moment du calibrage.

Cependant, l'utilisation d'un réseau de neurones lors de la calibration rend possible la spécification d'un modèle de choix discret basé sur une fonction d'utilité non paramétrique [10, 44, 46]. Cette approche permet de rendre compte d'une stochasticité complexe, à condition bien sûr que des données de calibration suffisamment riches soient disponibles. En contrepartie, comparativement à une approche paramétrique, la possibilité d'identifier et de donner un sens à certains attributs de la fonction d'utilité (comme dans notre cas le prix et la sensibilité au prix) est perdue.

Adoptant le point de vue de la théorie de l'utilité aléatoire, la désutilité d'une route r est représentée par la variable aléatoire

$$\tilde{u}_r = u_r + \tilde{\varepsilon}_r$$

où les termes d'erreur $\tilde{\varepsilon}_r \sim \text{Gumbel}(-\eta/\theta, \theta)$, $r \in \mathscr{R}$ sont iid (indépendants et identiquement distribués) sur l'ensemble des routes $\mathscr{R}$. Le terme η/θ est un paramètre de localisation, où η est la constante d'Euler, et θ est un paramètre d'échelle. On a

$$P\{\tilde{u}_r \leq \tilde{u}_\ell, \forall \ell\} = \exp(-\theta\, u_r) / \sum_{\ell \in \mathscr{R}} \exp(-\theta\, u_\ell).$$

L'affectation logit possède des propriétés attrayantes qu'il est utile de rappeler. Initialement, Dial a proposé une procédure efficace d'affectation des flots dans un réseau basée sur le modèle de choix logit. L'algorithme ne fait pas explicitement référence à l'ensemble des chemins [34], mais suppose que chaque chemin soit constitué d'arcs ne s'approchant pas (resp. ne s'éloignant pas) de l'origine (resp. de la destination). La complexité de cet algorithme se compare à celle d'un problème de plus court chemin. Van Vliet obtient l'expression des probabilités de transition entre deux sommets adjacents du réseau dans une analyse de la matrice de distribution des demandes sous-jacentes à une affectation multiproduits résultant de l'algorithme de Dial [77]. Akamatsu exprime plus généralement l'affectation logit comme une chaîne de Markov dans laquelle chaque probabilité de transition a elle-même une expression de type logit, et propose un algorithme permettant d'évaluer ces probabilités dans un réseau possiblement cyclique [2]. Baillon et Cominetti, dans un article traitant plus spécifiquement de réseaux congestionnés, mettent en place un cadre conceptuel relatif à une affectation markovienne dans lequel peuvent s'insérer à la fois l'affectation logit, l'affectation déterministe ou encore un modèle hybride. La prise de décision y est modélisée de façon indépendante à chaque sommet du graphe [6].

Comme le modèle déterministe, le modèle logit prête flanc à la critique. En effet, la propriété IIA, à la base de la dérivation de Luce [53], fait fi de la corrélation qui peut exister entre les chemins, bien qu'en pratique plusieurs d'entre eux auront inévitablement des arcs en commun. Des probabilités de choix contre-intuitives, voire paradoxales, peuvent résulter du fait que l'usager ne tient compte que d'attributs définis sur les chemins du réseau et a donc une perception biaisée de sa topologie. Nous faisons un bref survol de modèles de choix de chemins plus réalistes.

Selon le modèle probit, le vecteur de termes d'erreur $\tilde{\varepsilon}$ suit une distribution multinormale, dont la matrice de covariance reflète les corrélations induites par la topologie du réseau [79]. L'évaluation des probabilités de choix probit nécessite cependant l'usage de la simulation.

Le modèle logit est le plus simple représentant de la riche famille des modèles de choix GEV (*Generalized Extreme Value*, en anglais) [11, 48, 75]. Ces modèles ont l'avantage d'offrir des probabilités dont les expressions sont analytiques (quoique souvent fort complexes) et permettent de rendre compte de riches structures de corrélations. La corrélation entre les choix est modélisée par le biais d'une hiérarchisation de la prise de décision. Les probabilités de choix s'expriment alors sous la forme d'une fonction de probabilités conditionnelles et marginales ; probabilités qui font intervenir des termes d'erreur gumbeliens (mutuellement indépendants). Parmi les modèles GEV spécialement adaptés à la modélisation du choix de chemins, mentionnons les modèles logit combinatoire par paire (*Paired-Combinatorial Logit*, en anglais) et logit emboîté par arc (*Link-Nested Logit*, en anglais) [18, 61, 78].

Une autre approche (les modèles *PS-Logit* et *C-Logit*) consiste à intégrer un terme de correction à la désutilité espérée d'un chemin (c'est-à-dire le vecteur u) [9, 17]. Le modèle *C-Logit* modélise de cette manière la « visibilité » d'un chemin parmi l'ensemble de tous les chemins. Un chemin voit sa désutilité d'autant plus pénalisée qu'il partage une grande proportion de ses arcs avec d'autres chemins. La définition de la désutilité sous-jacente à cette approche fait intervenir des attributs qui ne sont pas sommés sur les arcs du réseau et il en résulte une affectation qui ne partage pas la propriété markovienne du modèle de choix logit. Une telle situation se présente aussi dans les modèles *PS-Logit* et logit combinatoire par paire.

Dans le cadre de cette thèse, nous nous contenterons de traiter du cas logit, le problème de tarification logit étant déja fortement NP-difficile. Notre analyse servira de premier pas vers d'autres modèles plus élaborés, par exemple ceux de la famille GEV.

1.3 Affectation multi-classes

Une modélisation multi-classes consiste en la spécification de modèles de demande particuliers à chacune des classes d'usagers sous-jacentes à une segmentation préalable de la population. Cette approche permet d'enrichir l'hétérogénéité du modèle et de rendre compte d'une stochasticité accrue de la loi de demande. Alors que la structure du problème de tarification demeure essentiellement la même avec un nombre fini de classes, la généralisation au cas multi-classes continu, dans laquelle l'appartenance à une classe est indiquée par un paramètre continu aléatoire, mène à un problème d'optimisation beaucoup plus difficile à résoudre et qui n'admet que rarement une expression analytique.

La valeur du temps, ou alternativement la sensibilité au prix, sont souvent utilisées pour spécifier une modélisation multi-classes d'un choix de route dans un réseau [56]. Sous une affectation déterministe, Marcotte et al. [56] développent des approximations qui permettent de trouver des solutions quasi-optimales à un problème de tarification multi-classes continu qui s'apparente au nôtre. L'approche de résolution adoptée dans notre troisième article, traitant du cas logit mixte (*Mixed-Logit* , en anglais), se présente comme une extension de ce travail.

Le modèle de choix discret logit mixte est une généralisation du modèle logit qui diffère substantiellement de celles qu'incarnent les autres modèles de la famille GEV. La fonction d'utilité sous-jacente se présente comme une somme pondérée d'attributs dans laquelle la pondération de chaque attribut est modélisée par une variable aléatoire. Il a été démontré que tout modèle de choix discret dérivé de la théorie de l'utilité aléatoire peut être exprimé sous la forme d'un logit mixte [75]. L'évaluation des probabilités logit mixte réclame généralement l'utilisation de simulations.

1.4 Le problème de tarification logit

L'ajout d'un terme aléatoire continu à la désutilité d'une route, que ce soit un terme d'erreur gumbelien (sous une affectation logit) ou une sensibilité au prix aléatoire (sous une affectation logit mixte ou multi-classes déterministe), mène à une formulation du problème de tarification qui peut être interprétée comme un lissage de la formulation déterministe. Ce lissage est comparable à celui obtenu en relaxant les contraintes de complémentarité par Dussault [36]. Notons cependant qu'aucune hypothèse comportementale n'est sous-jacente à ce lissage, qui sert d'artifice pour une mise en œuvre algorithmique.

Le problème de tarification logit prend la forme mathématique :

Programme 1.2.

$$\begin{aligned} \max_{t,x} \quad & \sum_{r \in \mathscr{R}} \exp[-\theta\, u_r(t)] \sum_{a \in r} t_a \Big/ \sum_{\ell \in \mathscr{R}} \exp[-\theta\, u_\ell(t)] \\ \text{sujet à} \quad & t_a = 0 \quad \text{si} \quad a \notin \mathscr{A}_{\text{toll}}. \end{aligned}$$

Le programme 1.2 est différentiable, non linéaire, en général non concave et ne possède pas de contraintes. La taille des problèmes envisagés rend difficile l'application de techniques d'optimisation globale telles que l'analyse par intervalles ou la programmation lipschitzienne [59, 60, 74]. Néanmoins, en exprimant la distribution logit comme la solution d'un problème d'entropie, il est possible de formuler le problème comme un programme bi-niveau et de faire ressortir la structure combinatoire sous-jacente. On obtient ainsi le programme

Programme 1.3.

$$\begin{aligned} \max_{t,x} \quad & \sum_{r \in \mathscr{R}} (x_r \sum_{a \in r} t_a) \\ \text{sujet à} \quad & t_a = 0 \quad \text{si} \quad a \notin \mathscr{A}_{\text{toll}} \\ & x \in \arg\min_y \left\{ \sum_{r \in \mathscr{R}} y_r \, u_r(t) + \frac{1}{\theta} y_r \log y_r : \sum_{r \in \mathscr{R}} y_r = 1, y \geq 0 \right\}, \end{aligned}$$

Dans les modèles ci-dessus, le paramètre d'échelle θ mesure la conformité des affectations déterministe et logit. Si θ est borné, le terme d'entropie joue le rôle d'une barrière à proximité de l'origine. Une valeur faible de θ procure une affectation quasi équiprobable des flots sur l'ensemble des chemins, un problème de tarification bien conditionné numériquement qui est peut-être même unimodal. Une valeur élevée de θ mène à un problème de second niveau s'apparentant à un problème de plus court chemin et où la plupart des composantes du vecteur x seront proches de zéro. Dans ce cas, le lagrangien du niveau inférieur sera mal conditionné et le Programme 1.3 sera difficile à résoudre. La formulation déterministe est obtenue comme le cas limite $\theta \to \infty$.

1.5 Approches de résolution

Cette thèse est composée de trois articles, chacun abordant le problème de tarification logit sous un angle différent. Le premier article traite des propriétés théoriques du modèle logit, dans le cas monoproduit. Nous y développons une analyse de sensibilité indépendante de la valeur que peut prendre le paramètre θ. En tirant profit de la nature markovienne de l'affectation logit, nous exprimons des conditions nécessaires d'optimalité qui font intervenir explicitement la topologie du réseau. Par l'introduction de notions appropriées d'équivalences entre réseaux, nous démontrons la validité de règles de simplification topologiques qui ne modifient pas la nature et la valeur des optima. Sur la base d'observations effectuées sur un strict sous-ensemble du réseau, nous montrons comment tirer des conclusions sur la structure d'un point stationnaire du problème originel. Ces règles permettent non seulement de réduire la taille des problèmes, mais aussi d'identifier une classe d'instances unimodales. Enfin, nous établissons un parallèle avec un problème de conception d'une ligne de produit, pour lequel nous généralisons certains de nos résultats.

La seconde publication traite du cas multi-produits. Nous y mettons en oeuvre des outils permettant d'identifier des solutions quasi-optimales, en utilisant diverses techniques algorithmiques : recherche locale initiée de plusieurs points de départ, dont la solution globale d'une approximation combinatoire du problème ; méthode de régions de confiance basée sur des approximations linéaires ou quadratiques. Les approximations combinatoires que nous avons obtenues permettent, étant donnée une puissance de calcul suffisante, d'approcher avec une précision arbitraire la fonction de revenu logit. Nous en avons profité pour obtenir des bornes supérieures sur l'erreur d'approximation associée à différents modèles combinatoires. La qualité de ces bornes a ensuite été validée par des tests numériques qui ont par ailleurs montré qu'il n'est pas en général nécessaire de faire une approximation très précise des probabilités de choix logit pour trouver les meilleures solutions. En effet, les solutions optimales des formulations logit et déterministes diffèrent beaucoup moins en termes des tarifs optimaux, bien qu'elles soient très différentes au niveau des flots et du revenu générés.

Dans le troisième article, nous considérons le cas plus général où la sensibilité au prix varie et pour lequel nous proposons une modélisation logit mixte. La fonction de revenu n'admet pas d'expression analytique ; nous définissons alors des approximations non linéaires et des approximations combinatoires. Comme dans le deuxième article, les approximations combinatoires sont utilisées pour identifier de « bonnes » régions de concavité et les approximations non linéaires pour implanter un algorithme de montée basé sur l'évaluation du gradient de l'objectif. Les tests numériques nous ont permis de constater que l'affectation probabiliste des flots et la modélisation multi-classes continue favorisaient les méthodes locales aux dépens d'algorithmes plus sophistiqués. En fait, un algorithme de montée que l'on démarre depuis un point choisi de façon aléatoire trouve de très bonnes solutions avec peu de tentatives, comparativement aux autres approches plus élaborées. Nous avons également étudié l'impact du paramètre de sensibilité au prix sur la provenance des revenus. Plus précisément, nous avons décrit de quelle manière les différents segments de la population (riche, pauvre, etc.) contribuaient le plus au revenu total par le biais de la « distribution de la contribution ». Cette fonction de densité donne, sous une tarification donnée, la probabilité qu'une unité de revenu soit générée par un usager dont la sensibilité au prix est située dans tel ou tel intervalle.

1.6 Problèmes connexes

Bien que peu d'articles aient traité de tarification logit sur des réseaux de transport dans le but de maximiser le profit d'une entreprise, certains sujets sont apparentés à nos travaux. Nous profitons de cette section pour les mentionner.

1.6.1 Tarification dans un réseau congestionné

Une affectation dans un réseau congestionné est en général modélisée en définissant la désutilité d'un arc par une fonction décroissante du flot qui y circule. Dans ce cadre, le premier principe de Warlop généralise l'affectation déterministe « tout-ou-rien » et stipule que seuls les plus courts chemins portent des flots positifs. Il en résulte une affectation qui disperse le flot et qui rappelle l'affectation logit, mais qui en fait est d'une nature totalement différente. Un équilibre usager déterministe « DUE » (*Deterministic User Equilibrium*, en anglais), est une affectation qui satisfait ce principe.

Bien que les modèles de base soient déterministes, il existe des extensions qui considèrent à la fois le phénomène de congestion et la stochasticité qu'impliquent les modèles de choix discret. L'affectation « SUE » (*Stochastic User Equilibrium*, en anglais) est une affectation logit dans un réseau congestionné. La propabilité P_r de choisir une route $r \in \mathscr{R}$ s'exprime alors

$$P_r = \exp[-\theta(u_r(t,P))] / \sum_{\ell \in \mathscr{R}} \exp[-\theta(u_\ell(t,P))],$$

où la désutilité $u_r(t,P)$ est une fonction décroissante de P_r.

Le problème de la tarification d'un réseau congestionné consiste à optimiser l'utilisation des ressources du réseau et appartient à l'une ou l'autre des deux catégories suivantes : la tarification *first-best*, pour laquelle des tarifs sont imposés sur tous les arcs du réseau, ou la tarification *second-best*, pour laquelle des tarifs sont imposés sur un strict sous-ensemble des arcs du réseau.

Dans le cas d'une tarification *first-best*, on peut démontrer, sous des hypothèses faibles, l'existence de tarifs qui induisent une répartition optimale du flot, même sous des modèles d'affectation multi-classes. Dans ce contexte, le problème perd largement sa nature combinatoire. En fait, dans bien des situations, il découle de résultats classiques en théorie économique que la tarification au coût marginal atteint le but cherché, ce qui a été démontré dans le cas de l'affectation stochastique SUE par Yang [80].

Différentes généralisations de l'affectation SUE à d'autres modèles de la famille GEV ont été introduites par le biais de la résolution de différents problèmes d'entropie généralisant le niveau inférieur du Programme 1.3. Par exemple, Prashker et Bekhor pour le modèle logit emboîté par arc dans [64] et Koppelman et Chen pour le modèle logit combinatoire par paires dans [49]. Toutefois, peu de travaux considèrent le problème de tarification *first-best* sur la base d'une affectation GEV plus élaborée que le modèle logit. Notons cependant l'article de Cominetti et Guzman [28] qui se penche sur un problème de tarification *first-best* dans un réseau de télécommunication, sous un équilibre usager stochastique de type markovien, et dans lequel un modèle de choix discret distinct est modélisé à chaque sommet du réseau.

Dans le cas *second-best*, le problème se formule comme un programme bi-niveau de grande taille auquel on peut appliquer une approche locale se basant sur l'analyse de sensibilité d'un équilibre par rapport aux tarifs (voir [62] dans le cas DUE). Le problème de la maximisation du revenu dans un cadre *second-best* est moins fréquemment abordé. Dans son mémoire de maîtrise, Poirier [63] considère une affectation DUE et utilise une formulation linéaire mixte dans laquelle la fonction de congestion est approximée par une fonction étagée. Dimitriou et al. proposent dans [35] un algorithme de montée sans dérivée sous une affectation SUE. Finalement, Sumalee et al. [72] considèrent le problème *second-best* sous un équilibre usager stochastique de type probit. La fonction objectif est exprimée avec beaucoup de généralité et admet la maximisation du revenu comme un cas particulier. Les auteurs proposent une analyse de sensibilité de la fonction objectif basée sur une approximation analytique des probabilités probit et présentent des résultats numériques obtenus avec un algorithme de montée sur des instances où une tarification uniforme est imposée aux arcs tarifés du réseau.

1.6.2 Tarification d'une ligne de produits

Le modèle logit apparaît fréquemment dans la littérature économique, en particulier dans la famille des problèmes de conception d'une ligne de produits [4, 5, 42, 45, 54, 67, 68]. Parmi ces problèmes, le problème de sélection et de tarification d'une ligne de produits s'apparente particulièrement au nôtre. Le problème de tarification d'une ligne de produits consiste, étant donné des coûts marginaux s_k, à identifier des prix t_k auxquels seront offerts les produits de la ligne $\mathscr{K}$, afin de maximiser les profits générés. Le problème de sélection d'une ligne de produits consiste à choisir un sous-ensemble de produits $\mathscr{K}' \subset \mathscr{K}$ à intégrer à la ligne. Sous une modélisation logit de la demande, le problème de sélection et de tarification d'une ligne de produit peut être exprimé

Programme 1.4.

$$\max_{\mathscr{K}' \subset 2^{\mathscr{K}}} \max_{t} \frac{\sum_{k \in \mathscr{K}'} \exp[-\theta(-c_k + t_k)]}{\sum_{k \in \mathscr{K}'} \exp[-\theta(-c_k + t_k)] + 1} (t_k - s_k).$$

où c_k est une mesure d'utilité à laquelle on se réfère sous le vocable de « prix de réserve » (*reservation price*, en anglais). La notion de surplus (*share of surplus*, en anglais) y joue un rôle analogue à celui de la désutilité pour le problème de tarification dans un réseau. Le surplus associé au produit k est donné par $-c_k + t_k$. Un produit nul dont le surplus est fixé à zéro joue ici un rôle analogue à celui d'un chemin de la compétition dans le problème de tarification dans un réseau. Les formulations adoptées dans [42, 45, 54, 67, 68] sont des spécialisations de ce cas de base.

Une approche alternative consiste à modéliser l'élasticité de la demande totale sur la base d'un surplus positif : une transaction est effectuée si et seulement si le prix d'au moins un des produits est inférieur à son prix de réserve. Dans ce cadre, voir Shioda [71] sous une modélisation déterministe de la demande et [5, 70] sous une modélisation stochastique.

Dans notre premier article, nous décrivons les liens qui unissent les problèmes de tarification d'un réseau et d'une ligne de produits. En particulier, nos résultats sur les réseaux parallèles peuvent être appliqués au problème de la tarification d'une ligne de produits, qui constitue, sous une modélisation logit de la demande avec produit nul, un cas particulier de notre problème. Réciproquement, certaines approches algorithmiques conçues pour les problèmes de conception d'une ligne de produits peuvent être adaptées au problème de tarification de réseau. Cependant, ces dernières n'exploitent généralement pas pleinement la nature combinatoire du problème, et notamment la topologie du réseau.

Hanson et Martin abordent le problème de la tarification d'une ligne de produits dans un contexte multi-classes et sur la base d'une fonction d'utilité exprimée avec beaucoup de généralité [42]. Les auteurs démontrent que le problème est concave pour une valeur suffisamment faible du paramètre d'échelle logit θ. Dans le cas d'une valeur arbitraire de θ, ils proposent un algorithme d'homotopie utilisant θ comme paramètre de lissage. Cette méthode n'est pas sans rappeler la méthode de lissage proposée par Dussault [36] pour le problème de tarification déterministe.

Certains auteurs [19, 67] abordent le problème de la sélection d'une ligne de produits dans lequel chaque produit est offert à une gamme de prix candidats, si bien que la sélection optimale procure une indication quand à la tarification optimale. Chen et Hausman [19] montrent que sous une modélisation logit d'une demande homogène, le problème est équivalent à un programme linéaire fractionnaire d'un seul terme et peut être résolu efficacement par une méthode standard [7]. Schon [67] considère le cas plus général d'une demande hétérogène et inclut des contraintes supplémentaires relatives aux coûts de production. Remarquons que des tarifs spécifiques sont imposés à chaque classe. Utilisant les résultats de Chen et Hausman, Schon montre comment résoudre le problème par le biais d'une séquence de sous-problèmes en nombres entiers.

Le problème de tarification et de sélection est à nouveau abordé par Schon sous une modélisation logit de la demande dans [68] avec, cette fois, des prix continus. Une formulation linéaire en nombres entiers est obtenue dans laquelle l'objectif est une fonction concave des probabilités de choix associées à chaque produit. Les variables de prix ne sont pas présentes dans cette formulation et les prix optimaux sont inférés sur la base des probabilités de choix optimales. Un produit est par ailleurs incorporé à la ligne si et seulement si la probabilité de choix associée est suffisamment positive. La possibilité, ici, de réexprimer le problème en termes des probabilités de choix uniquement, et le résultat de concavité sous-jacent, doivent être comparés à nos résultats d'unimodalité du problème de tarification défini sur des réseaux parallèles, que nous développons dans notre premier article.

1.6.3 Gestion du revenu

Le vocable « gestion du revenu » recouvre un ensemble de techniques permettant d'optimiser les profits d'une firme qui opère dans un milieu à forte concurrence et dont les coûts marginaux sont faibles en regard des coûts fixes. D'abord appliquée dans l'industrie aérienne, la pratique de la gestion du revenu s'est étendue aux domaines de l'hôtellerie, des croisières, du rail et des télécommunications.

Le modèle de gestion du revenu est situé dans un un cadre dynamique où l'horizon temporel est associé à l'avènement de l'expiration des produits. Sur chaque plage horaire, les variables de décision concernent la sélection et la tarification de l'ensemble de produits offert à ce moment. Une modélisation de type logit de la demande est adoptée dans [13, 73, 83]. Ce problème est à bien des égards une généralisation à un environnement dynamique du problème de sélection et tarification d'une ligne de produits. On peut par ailleurs représenter un choix de consommation par un parcours dans un réseau, où les contraintes associées au déploiement de l'horizon temporel sont modélisées par une topologie appropriée. Dès lors l'aspect « tarification » du problème de gestion des revenus se présente comme un cas particulier du problème de la tarification d'un réseau.

Vue la taille des instances généralement considérées, par exemple dans l'industrie aérienne, la tarification se fait de façon heuristique sur la base du « prix d'offre » (*bit-prices*, en anglais), une quantité qui procure un estimé ponctuel de l'impact des prix sur le revenu généré. Le prix d'offre est obtenu par le biais d'une approximation linéaire du problème dans lequel des contraintes de capacités sont associées aux différents produits et aux différentes plages horaires. Les multiplicateurs associés à ces contraintes sont appelés *bitprices*. Un multiplicateur positif suggère de monter un prix, un multiplicateur négatif, de le baisser [21].

CHAPITRE 2

LOGIT NETWORK PRICING

Dans le premier article, nous considérons le problème de tarification sur un réseau sous une affectation logit de la demande entre une origine et une destination. Nous présentons une analyse de premier ordre du problème qui tire profit des propriétés analytiques de la distribution logit et, notamment, de sa nature markovienne. Plusieurs de nos résultats découlent plus particulièrement de la propriété IIA, suivant laquelle les probabilités de choix de routes forment entre elles des proportions qui demeurent les mêmes suite au retrait où à l'ajout d'autres routes dans le réseau.

Soit $F(t)$ l'espérance de la taxe que paye un usager du réseau et $F(t|a)$ l'espérance conditionnelle de cette taxe étant donné que l'arc tarifé a est emprunté. Nous démontrons alors que pour un point critique t la relation

$$F(t \mid a) = F(t) + 1/\theta,$$

qui joue un rôle important dans notre analyse, est vérifiée. De cette identité nous concluons que la tarification optimale d'un réseau dont tous les arcs tarifés sont situés sur des routes distinctes (c'est-à-dire un « réseau parallèle ») impose une même taxe sur chacun d'entre eux. Il en résulte l'unimodalité du problème pour des réseaux parallèles, ce que nous démontrons par le biais d'un argument de monotonie faisant intervenir le système de Karush–Kuhn–Tucker.

À l'aide de notions d'équivalence entre réseaux, nous généralisons le résultat associé à l'identité ci-dessus et démontrons la validité de règles de simplification d'un réseau qui laissent les valeurs des maxima inchangées, permettant une réduction de la dimension du problème. Par exemple, un « goulot d'étranglement » est un arc tarifé qui apparaît sur chacune des routes tarifées. Nous démontrons qu'un vecteur de tarifs comportant uniquement une composante non nulle sur un goulot d'étranglement permet d'atteindre l'optimum. Ou encore une « coupe tarifée » est un ensemble d'arcs tarifés qui interceptent toutes les routes tarifées, chaque route tarifée utilisant exactement un arc dans cet ensemble. Nous démontrons qu'une tarification uniforme des arcs d'une coupe tarifée permet d'atteindre l'optimum. Sur la base de ces résultats, nous caractérisons une vaste gamme de réseaux qui sont, du point de vue de la tarification, équivalents à un réseau parallèle.

La notion de « cellule de réseau » nous permet de généraliser davantage nos résultats. Une cellule est un sous-ensemble d'arcs tarifés tel que i) pour accéder à la cellule, un usager doit d'abord visiter un sommet identifié comme le « point d'entrée » et ii) pour sortir de la cellule, un usager doit d'abord visiter un sommet identifié comme le « point de sortie ». En utilisant la propriété markovienne de l'affectation logit, nous démontrons que l'analyse d'une cellule du réseau permet de caractériser un point stationnaire par rapport au revenu généré dans l'ensemble du réseau. Plus précisément, les notions de goulot d'étranglement et de coupe tarifée, ainsi que les règles de simplification qui les accompagnent, peuvent être appliquées en se référant uniquement aux segments de parcours entre le point d'entrée et le point de sortie d'une cellule.

Finalement, nous faisons un parallèle avec le problème de tarification et de sélection d'une ligne de produits. Nous montrons que ce problème, dans sa forme la plus simple et sous une modélisation logit de la demande, est un cas particulier du nôtre. En particulier, notre analyse d'une tarification optimale dans un réseau parallèle se traduit, pour le problème de tarification d'une ligne de produits, en termes d'une tarification qui à l'optimalité rapporte une même marge de profit pour chaque produit offert. Bien que ce résultat ne soit pas neuf, la démonstration que nous en faisons se démarque par sa simplicité.

Submitted to Computers & Operations Research on December 23, 2011.

LOGIT NETWORK PRICING

FRANÇOIS GILBERT, PATRICE MARCOTTE AND GILLES SAVARD

ABSTRACT

We consider a toll setting problem defined over a transportation network where the underlying flow assignment is logit-based. This setting yields a nonlinear optimization problem that possesses strong combinatorial features. Taking advantage of the logit analytical properties, we analyze its first order optimality conditions with respect to the network topology. Simplification rules that leave the optima unchanged are introduced, and a class of unimodal instances is characterized. The connection with the more classical problem known as "product line pricing and selection" in economics is also emphasized.

Keywords: Network pricing, discrete choice models, product line design.

January 2012

2.1 Introduction

Let us consider the problem of setting optimal tolls on a subset of arcs of a single-commodity network, where commuters from the origin to the destination are assigned to paths minimizing a disutility function that takes into account fixed costs (lengths) and tolls. To provide a nontrivial account of commuters' relative awareness of the network conditions, we pose a logit probabilistic route choice model, i.e., path disutilities involve a random term that follows a Gumbel distribution. According to these assumptions, and in contrast with the "all-or-nothing" case, one obtains an assignment where flow is a nonlinear function of tolls.

In the context of route choice modeling, elaborate GEV (General Extreme Value) models such as the link-nested or pairwise combinatorial logit models offer more flexibility than the basic logit for modeling a decision process where correlated alternatives arise [18, 78]. However, this comes at the expense of complex probability expressions and involved calibration procedures, and motivates our choice of the logit model. Indeed, the analytical features of the latter make it attractive versus arguably more realistic models such as the probit or mixed logit, that require simulation or other approximation schemes for their evaluation [8, 23, 72, 75]. For instance, the Markovian property of the logit allows for the specification of a network loading procedure that avoids explicit path enumeration [2, 6, 34].

Our pricing problem consists in maximizing the expected revenue generated by the demand. It is related to choice-based revenue management [73], dynamic assortment problems [66] and product line design problems [5, 19, 42, 67, 68]. Actually the product line pricing problem, which under a logit modelling of the demand constitutes an instance of logit pricing that involves a particular network topology, will be fully reviewed in Section 4. It is also worth mentioning the connection with the problem of selecting optimal tolls within a stochastic user equilibrium environment (logit-based congested allocation) considered in [55, 80], although our focus is on profit maximization rather than network efficiency.

Until now, the network pricing problem has mainly been addressed under the assumption that commuter behavior is deterministic [12, 14, 65, 76], i.e., flow is only assigned to shortest paths, in contrast with a logit assignment where every path carries flow. We note that the deterministic formulation yields a combinatorial optimization problem that may be viewed as a zero variance limiting case of the logit based formulation. Conversely, the logit model can be considered a smooth extension of its deterministic counterpart. Since the deterministic network pricing problem is strongly NP-hard (see [65]), and has been solved to global optimality only for instances of relatively small size, it ensues that the logit pricing problem is both theoretically difficult and challenging from a computational point of view. It is a nonlinear and non-convex optimization problem that possesses several local optima, together with an underlying combinatorial structure inherited from the deterministic case.

In the view that the computation of a global optimum is "intractable" we will, in this paper, focus on analytical properties that yield insight into the structure of stationary solutions, and take explicitly into account the network topology. We will also provide rules that allow to simplify the network structure without modifying the nature of the problem, and consider a class of unimodal instances for which an optimal solution can easily be derived.

This paper is organized as follows. In Section 2.2, we describe the logit framework, together with its Markovian features, and formulate the logit pricing model. We also introduce the key notion of network equivalence and show how it can be used to simplify the problem. In Section 2.3, we investigate the properties of the toll setting problem defined over a network composed of uncorrelated paths. Parallel networks are considered in Section 2.4, where we make the connection with the product line design problem. In Section 2.5, we identify a class of unimodal instances and prove simplification rules for networks where the set of toll paths involves either *bottleneck tolls* or *parallel tolls*, the latter configuration having been defined in Section 2.3. In Section 2.6, we introduce the notion of *network cell*, which allows to generalize most of the previous results in a unified setting. Concluding remarks are given in Section 2.7.

2.2 The logit pricing model

Let us first consider the single-OD deterministic network pricing problem, which is best expressed, in path flow space, as the bilevel mathematical program:

Program 2.1.

$$
\begin{aligned}
\max_{t,x} \quad & \sum_{r\in\mathscr{R}} (x_r \sum_{a\in\mathscr{A}(r)} t_a) \\
s.t. \quad & t_a = 0 \;\; if \;\; a \notin \mathscr{A}_{\text{toll}} \\
& x \in \arg\min_{y} \left\{ \sum_r (y_r \sum_{a\in\mathscr{A}(r)} (c_a + t_a)) : \sum_r y_r = 1, y \geq 0 \right\},
\end{aligned}
$$

where $\mathscr{A}_{\text{toll}}$ denotes the set of toll arcs, t_a the toll on arc a, c_a the initial cost on arc a, x the set of user flows, and $\mathscr{R}$ the set of paths (routes). Also, we use non calligraphic letters to represent the cardinality of sets, i.e., $R = |\mathscr{R}|$ and $A_{\text{toll}} = |\mathscr{A}_{\text{toll}}|$. The network is assumed to be acyclic, and without loss of generality, demand is set to one. To avoid a trivial situation, we assume that $\mathscr{R}$ contains at least one path composed solely of toll-free arcs. For ease of reference, the main notations are displayed in the appendix.

The lower level of Program 2.1 insures that $u_r(t) = \sum_{a\in\mathscr{A}(r)}(c_a + t_a)$, referred to as path r's disutility, is minimal if and only if path r carries positive flow. The logit toll model generalizes Program 2.1 and assumes that these quantities are random. More precisely, let $\tilde{u}_r(t) = u_r(t) + \tilde{\varepsilon}_r$, where the $\tilde{\varepsilon}_r$'s are iid Gumbel variates. According to this assumption, the probability that a path r is of minimal disutility, i.e., $P_r^{\mathscr{R}} = \text{Prob}(\tilde{u}_r \leq \tilde{u}_{r'}, r' \in \mathscr{R})$ is given by the logit ratio:

$$
P_r^{\mathscr{R}}(t) = \frac{\exp[-\theta\, u_r(t)]}{\sum_{r'\in\mathscr{R}} \exp[-\theta\, u_{r'}(t)]}, \tag{2.1}
$$

where the scale parameter θ is inversely proportional to the standard deviation of the Gumbel random variable. For ease of notation, the upper index $\mathscr{R}$ will be dropped whenever no confusion can occur. The same remark applies to parameter t. Note that we have $P_r \to x_r$ as $\theta \to \infty$, where x is optimal for the lower level of Program 2.1. Also, we have $P_r = 1/R$, for $\theta = 0$. Our analysis will focus on the situation where the parameter θ is positive and finite.

Before introducing the logit pricing model, it is useful to consider some properties of the logit assignment. First, we recall that the decision process associated with the choice of a path can be equivalently expressed in terms of a sequence of decisions made at each node of the network, namely the choice of the next arc, until the destination is reached. Let a *path segment* from node i to node j denote a connected subset of the path's arcs, and $\tilde{\tau}_{ij}$ the corresponding disutility. The latter random variable follows a Gumbel distribution parameterized by t with expected value:

$$\tau_{ij}(t) = \mathbb{E}[\tilde{\tau}_{ij}(t)] = \mathbb{E}[\min\{\tilde{u}_r(t) | r \in \mathscr{R}^{ij}\}] = -\frac{1}{\theta} \log \sum_{r \in \mathscr{R}^{ij}} \exp[-\theta u_r(t)], \tag{2.2}$$

where $\mathscr{R}^{ij}$ is the set of path segments from node i to node j. Under this notation, we have that $P_r = \exp[-\theta(u_r - \tau_{od})]$. Now, let $\overline{P}_{ij}(t) = \exp[-\theta(u_{ij}(t) + \tau_{jd}(t) - \tau_{id}(t))]$ be the probability that a commuter located at node i select arc $(i, j) : j \in \mathscr{N}_i^+$, where $\mathscr{N}_i^+ = \{j \in \mathscr{N} : (i, j) \in \mathscr{A}\}$ denotes the set of successors of node i. We have:

$$\begin{aligned}
\prod_{i,j \in \mathscr{A}(r)} \overline{P}_{ij} &= \prod_{(i,j) \in \mathscr{A}(r)} \exp[-\theta(u_{ij} + \tau_{jd} - \tau_{id})] \\
&= \exp[-\theta \sum_{a \in \mathscr{A}(r)} (u_{ij} + \tau_{jd} - \tau_{id})] \\
&= \exp[-\theta(u_r - \tau_{od})] \\
&= P_r. \qquad (2.3)
\end{aligned}$$

Thus the matrix $\overline{P}$ describes a Markov chain whose transitions are logit-based. A network loading procedure that does not rely on the explicit enumeration of paths is constructed recursively as follows:

$$\tau_{ij} = \sum_{\ell \in \mathscr{N}_i^+} \overline{P}_{i\ell} (u_{i\ell} + \tau_{\ell j}), \tag{2.4}$$

where quantities τ and $\overline{P}$ are computed in a reverse topological order by Dial [34] given cycle free paths or using the reversibility of the transition matrix as done by Akamatsu [2] otherwise.

Most of what follows involves computations of quantities defined over the network's toll arcs. Let p denote the toll arc choice probability vector obtained by summing over paths incident to a given toll arc a, i.e.,

$$p_a(t) = \sum_{r\in\mathscr{R}} \delta_{ar} P_r(t),$$

where δ denotes the arc-path incidence matrix. Similarly, given the toll arcs a and b, let $p_{a\cap b}$ denote the probability of choosing a path incident to both a and b:

$$p_{a\cap b}(t) = \sum_{r\in\mathscr{R}} \delta_{ar}\delta_{br} P_r(t),$$

$p_{a|b}$ the conditional probability of choosing toll arc a given that toll arc b has been chosen:

$$p_{a|b}(t) = \sum_{r\in\mathscr{R}(b)} \delta_{ar} P_r^{\mathscr{R}(b)}(t)$$

where $\mathscr{R}(b) = \{r \in \mathscr{R} : b \in r\}$ is the set of paths in $\mathscr{R}$ incident to arc b. Finally, let $p_{a\cup b}$ the probability of choosing a path incident to either a or b:

$$p_{a\cup b}(t) = p_a(t) + p_b(t) - p_{a|b}(t).$$

The arc choice probability above is equivalently expressed as

$$p_a = \exp[-\theta(\tau_{o,a^-} + c_a + t_a + \tau_{a^+,d})],$$

where $a = (a^-, a^+)$. Even though we do make use of the Markovian property (2.3), it will be convenient to express our results in terms of the arc-path incidence matrix δ, or more specifically its restriction δ^{toll} to columns and rows associated with toll arcs and toll paths, respectively. Alternatively, we could have based our analysis on transition probabilities.

A key property of the logit model is the *independence of irrelevant alternatives* (IIA), which plays an important role in our analysis. It expresses the fact that *path elasticities*

$$E_a^r = \frac{\partial P_r}{\partial t_a}\frac{t_a}{P_r}.$$

are constant. Using $\partial \tau_{od}/\partial t_a = p_a$ we obtain

$$\begin{aligned}
\frac{\partial P_r}{\partial t_a} &= \frac{\partial \exp[-\theta(c_r+t_r)]}{\partial t_a}\exp(\theta\,\tau_{od}) + \exp[-\theta(c_r+t_r)]\frac{\partial \exp(\theta\,\tau_{od})}{\partial t_a} \\
&= -\theta(P_r\delta_{ar} + P_r\,p_a) \\
&= -\theta P_r\,(\delta_{ar} - p_a), \qquad (2.5)
\end{aligned}$$

and thus

$$E_a^r = -\theta\, t_a\,(\delta_{ar} - p_a). \qquad (2.6)$$

All paths, whether or not they are incident to a toll arc b, have identical elasticities with respect to t_b. In turn, path choice probabilities are affected by the same multiplicative constant after a change in the toll t_b. The immediate availability of sensitivity information for the logit model strongly contrasts with a probit or mixed logit assignment, which require simulation, or other approximation schemes, to be evaluated [72, 75].

It is possible to obtain a convenient expression for the partial derivatives of arc choice probabilities with respect to tolls. Let COV denote the toll arc choice covariance matrix, ie.,

$$\mathrm{COV}_{ab}(t) = p_{a\cap b}(t) - p_{a\cup b}(t).$$

We have:

$$\begin{aligned}
\partial p_a/\partial t_b &= -\theta\sum_r \delta_{ar}P_r(\delta_{br} - p_b) \\
&= -\theta\,\mathrm{COV}_{ab} \qquad (2.7)
\end{aligned}$$

and the corresponding elasticity expression $E_b^a = -\theta\, t_b\, \mathrm{COV}_{ab}/p_a$, where we define

$$E_b^a = \partial p_a/\partial t_b \cdot t_b/p_a.$$

The single commodity logit pricing problem is now expressed as

Program 2.2.

$$\begin{aligned}
&\max_t F(t) = p(t)\cdot t\\
&\textit{s.t.}\;\; p_a(t) = \sum_{r\in\mathscr{R}} \delta_{ra} \exp\left[-\theta\left(\sum_{b\in\mathscr{A}} \delta_{br} c_b + \sum_{b\in\mathscr{A}_{\text{toll}}} \delta_{br} t_b - \tau(t)\right)\right], \qquad a\in\mathscr{A}_{\text{toll}}\\
&\qquad \tau(t) = -\frac{1}{\theta}\log \sum_{r\in\mathscr{R}} \exp\left[-\theta\left(\sum_{b\in\mathscr{A}} \delta_{br} c_b + \sum_{b\in\mathscr{A}_{\text{toll}}} \delta_{br} t_b\right)\right].
\end{aligned}$$

In the above model, the scale parameter θ and the network topology are factors that impact the problem's difficulty. As θ gets small, the impact of costs on path disutilities becomes small, unless tolls are very large. If tolls are bounded from above, then these bounds will be tight at optimality, whenever the parameter θ is sufficiently small. On the other hand, as mentioned above, a large θ yields a more structured assignment with large choice probabilities assigned to shortest paths. This assignment is closer to the deterministic assignment, characterized by a "stiff" (in the numerical analysis sense) revenue curve, together with the presence of local maxima in the vicinity of each value of the toll vector t for which the shortest path is not unique. In the absence of specific assumptions on the network structure, one cannot expect a local method to yield the optimal solution of the logit pricing problem. Note that the formulation above is essentially unconstrained. Since it is differentiable, it is possible to implement a gradient-based algorithm for obtaining a solution that satisfies the first-order optimality conditions.

The Markovian property of the logit assignment and the proportional substitution patterns that it underlines will lead to network simplification rules that leaves optima essentially unchanged. In some way, this relates to simplifying rules introduced in the deterministic case and that allow to remove paths ("dominated paths") irrelevant to an optimal solution [12]. Note that, in contrast, paths cannot be ruled out *a priori* in the logit model since they all carry positive flow. To elaborate on this we need an appropriate notion of network equivalence.

Definition 1. *Two networks with expected revenue functions F^1 and F^2 and expected disutility τ^1 and τ^2 are* ***equivalent*** *if, for all t^1 (resp. t^2), there exists a t^2 (resp. t^1) such that $F^1(t^1) = F^2(t^2)$ and $\tau^1(t^1) = \tau^2(t^2)$.*

The following result shows that paths incident to the same set of toll arcs need not be distinguished. In some sense, this generalizes the notion of dominance mentioned above.

Theorem 2.1. *Any logit pricing problem is equivalent to a problem defined over a network where every subset of toll arcs is incident to at most one path.*

Proof. The result is shown by induction on the number of paths incident to identical sets of toll arcs. The theorem is trivially verified for the base case, that is, where all sets of paths incident to the same toll arcs have cardinality 1.

Consider a network $\mathscr{R}$ and its associated set of paths. Assume that there exists a subset $\mathscr{R}' \subset \mathscr{R}$ such that all paths in $\mathscr{R}'$ are incident to the same toll arcs and no other with $R' > 1$. Consider the following network with associate set of paths $\mathscr{R}^* = \{r'\} \cup \mathscr{R} \setminus \mathscr{R}'$, where r' is a toll path such that $c_{r'} = -\theta^{-1} \log \sum_{r \in \mathscr{R}'} \exp[-\theta(c_r)]$. Let $t \in \mathbb{R}^{R_{\text{toll}}}$ and $t' \in \mathbb{R}^{R^*_{\text{toll}}}$, where t is feasible in the first network and t^* is such that $r \in \mathscr{R} \setminus \mathscr{R}' \Rightarrow t^*_r = t_r$ and $t^*_{r'} = k$ where $r \in \mathscr{R}' \Rightarrow t_r = k$. By construction t^* is feasible for the second network, $r \in \mathscr{R} \setminus \mathscr{R}' \Rightarrow P_r^{\mathscr{R}}(t) = P_r^{\mathscr{R}^*}(t^*)$ and $\sum_{r \in \mathscr{R}'} P_r^{\mathscr{R}}(t) = P_{r'}^{\mathscr{R}^*}(t^*)$. Let $F^{\mathscr{R}}$ (respectively $F^{\mathscr{R}^*}$) the revenue generated on the network defined with reference to the set of path $\mathscr{R}$ (resp. $\mathscr{R}^*$). We have that

$$\begin{aligned} F^{\mathscr{R}}(t) &= \sum_{r\in\mathscr{R}'} P_r^{\mathscr{R}}(t)t_r + \sum_{r\in\mathscr{R}\setminus\mathscr{R}'} P_r^{\mathscr{R}}(t)t_r \\ &= \sum_{r\in\mathscr{R}'} P_r^{\mathscr{R}}(t)k + \sum_{r\in\mathscr{R}\setminus\mathscr{R}'} P_r^{\mathscr{R}}(t)t_r \\ &= P_{r'}^{\mathscr{R}^*}(t^*)t_{r'}^* + \sum_{r\in\mathscr{R}\setminus\mathscr{R}'} P_r^{\mathscr{R}^*}(t^*)t_r^* \\ &= F^{\mathscr{R}^*}(t^*). \end{aligned}$$

The result follows by induction, repeating the argument for each $\mathscr{R}'' \subset \mathscr{R}^*$ such that all paths in $\mathscr{R}''$ are incident to the same toll arcs and no other with $R'' > 1$. □

2.3 Parallel networks

In this section, we address pricing problems defined over a specific network topology. More precisely, we analyze the first order optimality conditions of Program 2.2 for *parallel networks*, in the sense of the following definition.

Definition 2. *A network is* **parallel** *if for any toll arcs a and b and any toll vector t, the conditional probability $p_{a|b}(t)$ only takes values zero or one.*

Examples of parallel networks are shown in Figure 1. In the simplest case, a parallel network contains a single toll arc, and we will show that in this case the pricing problem admits a single optimum. This conclusion clearly holds in the deterministic case, where the optimal strategy consists in setting the toll so that both the toll path and the shortest toll-free path have equal disutilities. However, the situation is less clear under the probabilistic model where, in the absence of an explicit expression for the solution, some form of argument is required.

Theorem 2.2. *Program 2.2 defined over a network that contains a single toll arc is unimodal.*

Proof. To prove the result, we show that there exists a single point that satisfies the first order conditions, i.e., where the gradient is equal to zero.

From Theorem 2.1, the network can be assumed to contain a single toll path and a single competition path. Since path choice probabilities are not affected by the addition of the same constant to all path disutilities, and given $c^* = c + ek$ we have $P_r \propto \exp[-\theta(c_r + t_r)] = \exp(\theta\, k)\, \exp[-\theta(c^*_r + t_r)] \propto \exp[-\theta(c^*_r + t_r)]$, we may set, without lost of generality, the disutility of the competition path to zero. Let c be the fixed cost of the toll path, and let $P(t)$ be the associated logit probability. Then $F(t) = P(t)\, t$ and, based on (2.5),

$$F'(t) = P(t) - \theta P(t)[1 - P(t)]\, t,$$

from which we infer

$$\begin{aligned} F'(t) = 0 \quad &\Leftrightarrow \quad 1 - \theta[1 - P(t)]\, t = 0 \\ &\Leftrightarrow \quad h(t) \equiv 1 - \theta\, t + \exp[-\theta(c + t)] = 0. \end{aligned}$$

The fact that $h(0) > 0$ and $\lim_{t\to\infty} h(t) < 0$ implies the existence of a first order point. Next $h'(t) = -\theta(1 + \exp[-\theta(c + t)]) < 0$ for every t implies strict monotonicity of h, hence the uniqueness of the solution. □

Corollary 2.1. *In a network containing a single toll arc, the revenue function is strictly pseudoconcave.*

In the general case we have the following characterization of a first order point.

Theorem 2.3. *Let $F(t|a) = \sum_b p_{b|a}(t)\, t_b$ denote the conditional expectation of the toll levied, given that a path incident to toll arc a is selected. Then t is a first order point of Program 2.2 if and only if*

$$F(t|a) = F(t) + 1/\theta, \quad \forall a \in \mathscr{A}_{\text{toll}}.$$

Equivalently we have $\sum_{b\in\mathscr{A}_{\text{toll}}} E^a_b = 1/\theta,\ \forall a \in \mathscr{A}_{\text{toll}}$.

Proof. Based on (2.7) we have $\nabla F(t) = p(t) - \theta \, \mathrm{COV}(t)\, t = 0$. Thus

$$\nabla_a F(t) = p_a(t)(1 - \theta \sum_b [p_{b|a}(t) - p_b(t)]\, t_b) = 0 \Leftrightarrow F(t|a) = F(t) + 1/\theta.$$

Using (2.6) we infer

$$\sum_{b \in \mathscr{A}_{\text{toll}}} E_b^a = \sum_{b \in \mathscr{A}_{\text{toll}}} (p_{b|a} - p_b)\, t_b = F(t|b) - F(t) = 1/\theta.$$

□

The above result provides information on the structure of a local optimum in terms that involves the network topology. It namely implies the following result.

Corollary 2.2. *If t is a first order point of Program 2.2, then $F(t|a) = F(t|b), \forall a, b \in \mathscr{A}_{\text{toll}}$.*

In parallel networks one may assume, without loss of generality, that arcs connect directly the origin and the destination. Arc and path choice probabilities then coincide, and toll arc choice elasticities are constant: $b \in \mathscr{A}_{\text{toll}} \Rightarrow E_b^a = \theta p_b, \forall a \in \mathscr{A}_{\text{toll}}$. An optimal solution of Program 2.2 defined over a parallel network has the very simple structure specified below.

Theorem 2.4. *Over a parallel network, Program 2.2 is unimodal and, at optimality, all tolls are equal.*

Proof. Let Network A be a parallel network with fixed cost vector c^A, where $\dim(c^A) = n+1$. Assume without lost of generality a single toll free path of fixed cost c_0^A. Let Network B be composed of two arcs connecting directly the origin and the destination, and be endowed with the fixed cost vector c^B. Assume without loss of generality and on account of Theorem 2.1 a single toll free path of fixed cost $c_0^B = c_0^A$. Furthermore, let the toll path's fixed cost be $c_1^B = -\theta^{-1} \log \sum_{r=1}^n \exp(-\theta c_r^A)$. We will show that (i) the pricing problem defined over Network A is a relaxation of that defined over Network B; (ii) given that all tolls in Network A are set to some value k and that the toll of Network

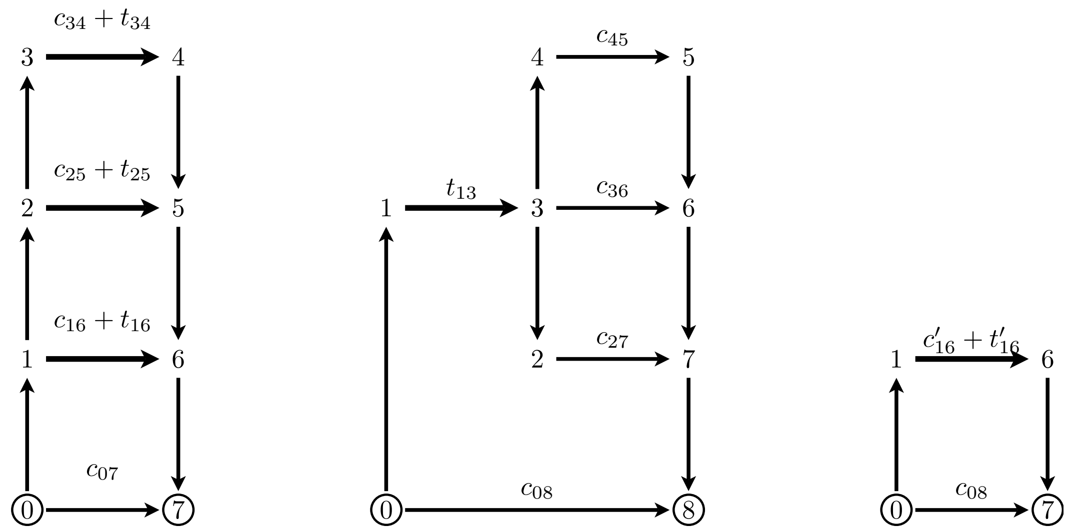


Figure 2.1: Three parallel networks. The two leftmost networks are first order equivalent by Theorem 2.4. The two rightmost networks are equivalent by Theorem 2.1.

B is also set to k, both networks generate an identical revenue; (iii) identical tolls are set at the optimum of Network A. Since identical tolls are set on each arc of Network A, all its toll arcs can be replaced by a single toll arc, shared by all toll paths. Following Theorem 2.1, the remaining toll free path segments can be replaced by a single toll free arc of fixed cost c_1^B. This proves (i) and (ii). Next, let t^* denote a first order point for Network A. From Corollary 2.2:

$$t_a^* = F(t^*|a) = F(t^*|b) = t_b^* \text{ for all } a, b \in \mathscr{A}_{\text{toll}}$$

and identical tolls are set at the optimum of Network A. This shows (iii) and the conclusion follows. □

An illustration of the above result is provided in Figure 2.1, where all three networks shown therein are parallel. The middle network generates the same revenue as the one on the left, given that $t_{34} = t_{25} = t_{16} = t_{13}$ and, according to Theorem 2.1, the middle network is equivalent to the one on the right given that

$$\exp(-\theta c'_{16}) = \exp(-\theta c_{45}) + \exp(-\theta c_{36}) + \exp(-\theta c_{27}).$$

Theorem 2.4 states that, at the optimum of Program 2.2 defined over a parallel network, a simple transformation of the network topology leaves the revenue unchanged, and the network becomes equivalent to a two-path single toll network. This prompts us to adapt the notion of equivalent network accordingly:

Definition 3. *Two networks with expected revenue functions F^1 and F^2 and expected disutility functions τ^1 and τ^2 are* ***first order equivalent*** *if, for all t^1 (resp. t^2) such that $\nabla F^1(t^1) = 0$ and $\nabla F^2(t^2) = 0$, there exists a toll t^2 (resp. t^1) such that $F^1(t^1) = F^2(t^2)$ and $\tau^1(t^1) = \tau^2(t^2)$.*

According to this definition, a parallel network is first order equivalent to a single toll network, as illustrated in Figure 2.1.

2.4 The product line problem

There is a close relationship between the network design problems and the product-line design problems. The latter consists in maximizing the revenue of a reseller offering a given set of products to utility-maximizing customers. Associated decisions concern the design and/or selection of products, the setting of their prices, as well as inventory related decisions. Buyer maximizes their *share of surplus*, expressed as the difference between each product's price and the buyer's corresponding *reservation price*, a measure of valuation that parallels a path's fixed cost in the network pricing environment. In most deterministic settings, a buyer selects a product of maximal positive surplus [32, 43, 50]. From a network pricing perspective, this is equivalent to setting the smallest toll free fixed cost to zero. A probabilistic approach proposed in [70] assigns demand to products with positive surplus proportionally to the ratio of their reservation prices. Note that such a ratio does not involve product prices and is thus only a function of the number of products associated with a positive surplus. Similar to the deterministic case, this yields a mixed integer optimization problem. Aydin generalizes the share of surplus framework to a logit behavioral model in [5]. Reservation prices are treated as Gumbel variates, which allows to derive a closed form probability expression for the probability of a positive surplus and, conditioned on this, for the choice of a specific product. This approach is reminiscent of the nested logit hierarchical choice model adopted in [4, 52], and that is cast within an oligopoly framework.

Most logit approaches to the product line pricing problem introduce a null product with a zero surplus (no-buy option) [19, 42, 45, 54, 67, 68]. While this can still be compared to a competition path, from a network pricing perspective, the corresponding probability does not model a non positive surplus. To the best of our knowledge, the positive surplus framework under a probabilistic modeling of commuter behavior has only been considered in [5, 70]. Other approaches to modeling elasticity with regard to the total demand are similar to that adopted in our model.

In the following, although we are primarily concerned with the pricing problem, we will also make the connection with the problem of jointly selecting and pricing a product line. Whenever possible, we will keep to our previous notation. While this might seem unnatural from a product line perspective, it does make the exposition more concise and allows to draw parallels with the network situation.

Let us consider a product line problem involving K products, each product k being endowed with a reservation price c_k, a price (to be determined) t_k, and a procurement cost s_k. We denote by $p_k(t)$ the probability that product k is purchased. In the absence of additional production fixed costs or resource related constraints, the logit-based product line joint selection and pricing problem is formulated as

Program 2.3.

$$\begin{aligned} \max_{\mathscr{K}' \subset 2^{\mathscr{K}}} \max_{t} \quad G_{\mathscr{K}'}(t) &= \sum_{k \in \mathscr{K}'} p_k(t)\,(t_k - s_k) \\ s.t. \qquad p_k(t) &= \frac{\exp[-\theta(-c_k + t_k)]}{\sum_{k \in \mathscr{P}} \exp[-\theta(-c_k + t_k)] + 1} \qquad k \in \mathscr{K}'. \end{aligned}$$

Let us first consider the pricing problem for a given set of products $\mathscr{K} = [1, \ldots, K]$.

Theorem 2.5. *Let $t \in \arg\max G_{\mathscr{K}}(\cdot)$ for a fixed set of products $\mathscr{K}$. Then there exists an optimal profit margin v such that $t_k = v + s_k, \forall k \in \mathscr{K}$, i.e, the profit margins of all products are equal at the optimum.*

Proof. Based on (2.7) we obtain

$$\nabla G_{\mathscr{K}}(t) = p(t) - \theta\, \mathrm{COV}(t)\,(t - s).$$

Next, invoking arguments similar to those used in the proof of Theorem 2.3, we obtain

$$\nabla_{t_k} G_{\mathscr{K}}(t) = 0 \;\Leftrightarrow\; G_{\mathscr{K}}(t|k) = G(t) + 1/\theta,$$

where $G_{\mathscr{K}}(t|k)$ denotes the conditional expectation of the profit, given that product k is purchased. The result then follows from the equality of $G(t|k)$ and $t_k - s_k$. □

Although similar results have been obtained in other settings, for instance in [45] within a Bayesian framework to account for a random brand effect; in [5] within a share of surplus framework; in [4] within a multifirm environment - the simplicity of the proof above is in stark contrast, and will serve as a basis for results pertaining to general network topologies.

Let us now consider the joint selection and pricing problem. Since logit probabilities are positive under any finite values of its parameters, setting the prices of any subset of the products to an arbitrary large value yields arbitrarily small corresponding choice probabilities. It is equivalent, in effect, to removing the associated products from the line. However, doing so is suboptimal.

Theorem 2.6. *At the unique optimal solution of Program 2.3, all products are selected.*

Proof. Let $\mathscr{K}' \subset 2^{\mathscr{K}}$ be a subset of $\mathscr{K}$ and

$$A = \sum_{k \in \mathscr{K}'} \exp[-\theta(-c_k + s_k)],$$

where $\mathcal{K}' \subset 2^{\mathcal{K}}$. Trivially, A reaches its maximal value when $\mathcal{K}' = 2^{\mathcal{K}}$. Now, let $H(v) = G(ev+s)$, where $e = [1,\dots,1]^T$, the profit function expressed as a function of the profit margin, given that these margins are equal for all products offered. A simple computation yields

$$H(v) \quad = \quad \frac{vA}{A+\exp(\theta v)}.$$

Since this function is increasing in A, it follows that it is optimal to offer all products. Next, the maximum of H is achieved at a point where the derivative is zero, i.e.,

$$A = (\theta v - 1)\exp(\theta v).$$

Since the function on the right-hand side is strictly increasing from a negative value -1 when $v = 0$ to $+\infty$, the equation has a unique solution and the result follows. □

In the presence of additional constraints, such as upper bounds on the toll vector, the above result may fail to hold, and one might have to resort to enumeration to solve the selection problem. For a given price vector, the optimal selection is made of products associated with *maximal* profit margins (see [73]). However, even for fixed prices, it has been shown in [66] that the maximal profit margin rule does not hold any more if the number of products must not exceed some predetermined number, although a polynomial algorithm for determining the optimal subset is available. If the subset of products is fixed, a path following method for solving a pricing problem involving several customer classes, as well as a very general utility function framework, has been proposed in [42]. The starting point of the homotopy corresponds to setting the parameter θ to a sufficiently small value. This models buyers as price insensitive and results in a pricing problem that is unimodal and easy to solve. The homotopy path then corresponds to increasing θ up to its true value.

A formulation of the joint selection and pricing problem involving resource related constraints is given in [68]. For small instances, it can be solved to global optimality by commercial nonlinear mixed integer solvers. Again, the underlying idea is that, since arbitrarily large tolls are allowed, the optimal prices obtained above yield the optimal selection. In contrast, on account of the additional constraints, the optimal selection issue is nontrivial.

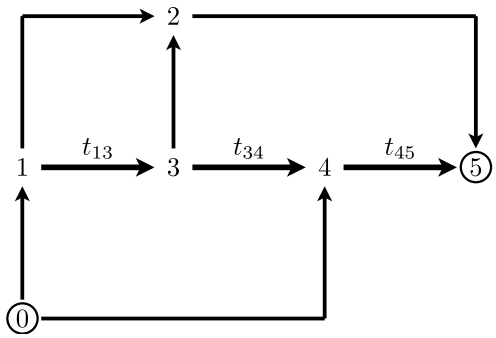
When the operation consisting in inferring the optimal selection from the optimal prices is not available, the only approach to the joint selection and pricing problem that we are aware of consists in solving a restriction of the problem to a set of discrete candidate tolls. This yields a linear fractional problem (assuming either a single class of potential buyers or class specific prices) that can be addressed efficiently on account of its specific structure [19, 67]. A comment is in order. Candidate tolls should be selected so as to yield equal profit margins. This is tantamount to replacing the set of candidate tolls by a set of candidate marginal profits. Indeed it is easy to show that optimal prices are such that the equal profit margin rule holds when prices lie within the interior of the feasible set.

2.5 A class of unimodal instances

In this section, we propose sufficient conditions for a network to be first order equivalent to a parallel network, thus defining a class of unimodal instances. More precisely, let us consider the following relaxation of Program 2.2, which removes all correlations between paths.

Definition 4. *A* ***parallel relaxation*** *B of Network A is obtained in the following way. To every path in A corresponds an arc in B connecting the origin and the destination directly, with matching costs. Similarly, to each toll path in A corresponds a toll arc in B.*

Note that, if arbitrary path tolls can be induced by some assignment of arc tolls, then the network is equivalent to its parallel relaxation. If a network topology is such that, for any given scalar $k \in \mathbb{R}$, appropriate arc tolls yield equal path tolls of value k, then the corresponding network is first order equivalent to its parallel relaxation. In either case, solving the parallel relaxation yields a global optimum of the initial problem.



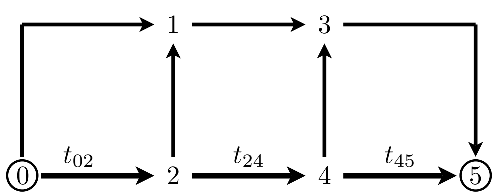
	(1,3)	(3,4)	(4,5)
r_1	1		
r_2	1	1	1
r_3			1

$$r_1 = (0,1,3,2,5)$$
$$r_2 = (0,1,3,4,5)$$
$$r_3 = (0,4,5)$$

Figure 2.2: Incidence matrix δ^{toll} has full row-rank (top right). The network is equivalent to a parallel network. Setting identical path tolls yields $t_{13} = -t_{34} = t_{45}$.

Theorem 2.7. *Let δ^{toll}, the restriction of the path-arc incidence matrix δ to the set of columns and rows associated with toll arcs and toll paths, respectively, be such that, for all $k > 0$, there exists t such that $\delta^{\text{toll}} t = ke$, where $e = [1, \cdots, 1]^T$. Then Program 2.2 is unimodal.*

Proof. By construction, arc tolls can be set so that all path tolls share a common value k, whatever k. By Corollary 2.2, at the optimum of the associated relaxed parallel network, all tolls are identical to say t^*. The conclusion follows after setting k to t^*. □

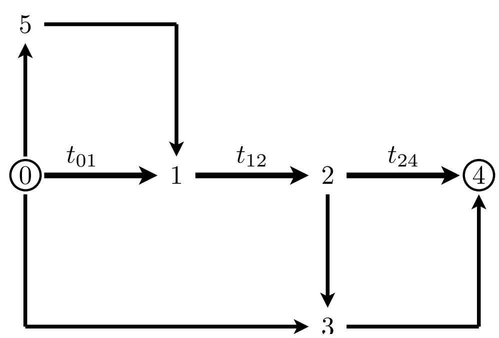


	(0,2)	(2,4)	(4,5)
r_1	1		
r_2	1	1	
r_3	1	1	1

$$r_1 = (0,2,1,3,5)$$
$$r_2 = (0,2,4,3,5)$$
$$r_3 = (0,2,4,5)$$

Figure 2.3: First order equivalent parallel network (δ^{toll} has full row-rank). Setting identical path tolls yields $t_{24} = t_{45} = 0$. Toll t_{02} forms a bottleneck.

The result is illustrated in Figures 2.2 through 2.6. Note that, given that δ^{toll} has full row-rank, the assumption of Theorem 2.7 is satisfied. This is the case in the network of Figure 2.2, and the optimum is found by solving its parallel relaxation. This yields a unimodal optimization problem whose optimum lies on the line $t_{13} = -t_{34} = t_{23}$. The networks of Figure 2.3 and Figure 2.4 also satisfy the assumptions of Theorem 2.7. In each case, a unique toll assumes a nonzero value at the optimum, and it is incident to all toll paths. Such toll arc constitutes a **bottleneck**. Only bottleneck tolls need be nonzero, and the associated network is first order equivalent to a parallel network. The situation differs somewhat for the networks of Figure 2.5 and Figure 2.6. Even though no single toll arc acts as a bottleneck, a strict subset of the toll arcs forms a cut of the entire set of toll paths, i.e., every toll path is incident to exactly one toll arc in the **cut**. At the optimum, only the tolls in the cut need to take (identical) nonzero values. Obviously, the notion of cut generalizes that of bottleneck, and a network that involves a cut is first order equivalent to a parallel network, which in turn is first order equivalent to a network with a bottleneck.



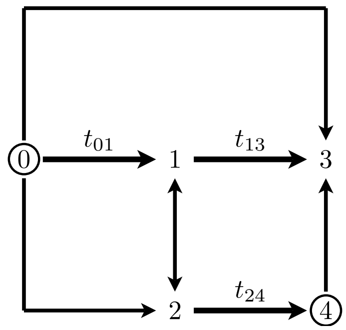
	(0,1)	(1,2)	(2,4)
r_1	1	1	1
r_2	1	1	
r_3		1	1
r_4		1	

$$r_1 = (0,1,2,4)$$
$$r_2 = (0,1,2,3,4)$$
$$r_3 = (0,5,1,2,4)$$
$$r_4 = (0,5,1,2,3,4)$$

Figure 2.4: First order equivalent parallel network. Setting identical path tolls yields $t_{0,1} = t_{2,4} = 0$. Toll $t_{1,2}$ forms a bottleneck.

2.6 Network cells

The notion of network cell that we now introduce allows for a generalization of the previous equivalence results to arbitrary topologies, by focusing on a strict subset of the network.



	(0,1)	(1,3)	(2,4)
r_1	1	1	
r_2	1		1
r_3		1	
r_4			1

$$\begin{aligned} r_1 &= (0,1,3) \\ r_2 &= (0,1,2,4,3) \\ r_3 &= (0,2,1,3) \\ r_4 &= (0,2,4,3) \end{aligned}$$

Figure 2.5: First order equivalent parallel network. Setting identical path tolls yields $t_{1,3} = t_{2,4}$ and $t_{0,1} = 0$. Tolls $t_{1,3}$ and $t_{2,4}$ form a toll cut.

Definition 5. *Given a graph $\mathscr{G} = (\mathscr{N}, \mathscr{A})$, an* ***entry point*** *and an* ***exit point*** $o', d' \in \mathscr{N}$*, respectively, a* **network cell** *is a triple $(o', d', \mathscr{A}')$ with $\mathscr{A}' \subset \mathscr{A}_{\text{toll}}$ such that:*

1. *path segments starting at a toll arc $a \notin \mathscr{A}'$ and ending at a toll arc $b \in \mathscr{A}'$ pass through the entry point;*

2. *path segments starting at a toll arc $a \in \mathscr{A}'$ and ending at a toll arc $b \notin \mathscr{A}'$ pass through the exit point.*

In terms of elasticities, this translates into $E_b^{a_1} = E_b^{a_2}$ for $a_1, a_2 \in \mathscr{A}'$ and $b \notin \mathscr{A}'$. Using the Markovian property (2.3) of the logit assignment we obtain the following generalization of Theorem 3.

Theorem 2.8. *Let t be a first order point of Program 2.2 defined over a network with associated graph $\mathscr{G} = (\mathscr{N}, \mathscr{A})$. For every $a \in \mathscr{A}'$ such that $\mathscr{G}' = (o', d', \mathscr{A}')$ is a cell we have:*

$$F^{\mathscr{A}'}(t|a) = F(t) - F^{\mathscr{A} - \mathscr{A}'}(t|\mathscr{A}') + 1/\theta \,,$$

where $F^X(t|Y) = \sum_{a \in X} p_{a|Y} t_a$ and $p_{a|Y}$ denotes the conditional probability of selecting a path incident to arc a, given that a path incident to at least one of the arcs in Y is

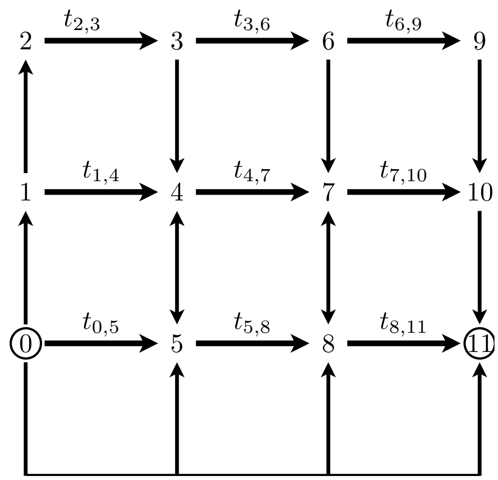


Figure 2.6: First order equivalent parallel network. Setting identical path tolls yields $t_{69} = t_{7,10} = t_{8,11}$ and all other tolls zero. These three arcs form a toll cut.

selected, i.e.,

$$p_{a|Y} = \sum_{r \in \mathscr{R}(Y)} \delta_{ar} P_r^{\mathscr{R}(Y)}$$

with $\mathscr{R}(Y) = \{r \in \mathscr{R} | r \cap Y \neq \emptyset\}$.

Proof. Distinguishing the revenue generated inside or outside the cell we have:

$$F(t|a) = F^{\mathscr{A}'}(t|a) + F^{\mathscr{A}-\mathscr{A}'}(t|a).$$

To prove the result it suffices to show that $p_{b|a'} = p_{b|\mathscr{A}'}$ for all $b \notin \mathscr{A}'$. Either arc b appears in a path incident to the cell or it does not. In the latter case we have $p_{b|a'} = p_{b|\mathscr{A}'} = 0$. Otherwise, since the sequence of nodes visited by a commuter forms a Markov chain, we obtain

$$p_{b|a'} = p_{b|o'} p_{o'|a'} = p_{b|o'} = p_{b|\mathscr{A}'}.$$

□

Note that for a cell $(o', d', \mathscr{A}')$, given that $a \in \mathscr{A}'$, the quantity $F^{\mathscr{A}'}(t|a)$ is a function of the tolls set in $\mathscr{A}'$ alone, and is therefore independent of the toll policy applied to the rest of the network. Thus information about a first order point can be inferred from the cell information alone, that is, in several situations, from a strict subset of the network's graph. This statement can be rewritten in the following way.

Corollary 2.3. *Let t be a first order point of Program 2.2 and $(o', d', \mathscr{A}')$ a network cell. Then, for all $a, b \in \mathscr{A}'$, the following holds:*

$$F^{\mathscr{A}'}(t|a) = F^{\mathscr{A}'}(t|b).$$

Our final result generalizes Theorem 2.7:

Theorem 2.9. *Let $\mathscr{G}$ be a graph with a cell $(o', d', \mathscr{A}'_{\text{toll}})$, and δ'_{toll} the restriction of δ^{toll} to the columns and rows associated with the arcs in $\mathscr{A}'$ and the paths incident to them. Assume that, for all $k > 0$, there exists $t \in \mathbb{R}^{A'}$ such that $\delta'_{\text{toll}} t = ke$. Then there exists a network that is first order equivalent to $\mathscr{G}$ where all toll arcs in $\mathscr{A}'$ are replaced by a single toll arc connecting node o' and node d'.*

Proof. Consider the restriction of the pricing problem to the set of path segments from o' to d', that is, $\max F^{\mathscr{A}'}(t|\mathscr{A}'')$ where $\mathscr{A}''$ is the set of all arcs appearing in a path segment between entry point o' and exit point d'. By Corollary 2.2, the hypothesis above ensures that (i) solving the parallel relaxation provides the optimal tolls; (ii) identical tolls are set on each path segment from o' to d'; (iii) all toll paths from o' to d' can be replaced by a single toll arc, providing a first order equivalent network. Corollary 2.3 insures that (ii) and (iii) remains verified at the optimum of the toll setting problem defined over the entire network. □

The simplification procedure described above is carried out in this last example by showing how the notions of parallel and bottleneck tolls are applied to network cells. Consider the network of Figure 2.7 and the three cells:

$$\begin{aligned}\text{cell}_1 &= (\{(3,4),(2,4)\},1,4),\\ \text{cell}_2 &= (\{(4,6),(4,5)\},4,7),\\ \text{cell}_3 &= (\{(9,10),(3,10),(10,11),(10,6)\},3,6).\end{aligned}$$

The toll arcs in cell$_1$ and cell$_2$ are set up in parallel and are thus identically tolled: $t_{34} = t_{24}$ in cell$_1$ and $t_{46} = t_{45}$ in cell$_2$. In cell$_3$ toll arcs $(9,10)$ and $(3,10)$ form a cut and thus $t_{9,10} = t_{3,10}$ and $t_{10,11} = t_{10,6} = 0$.

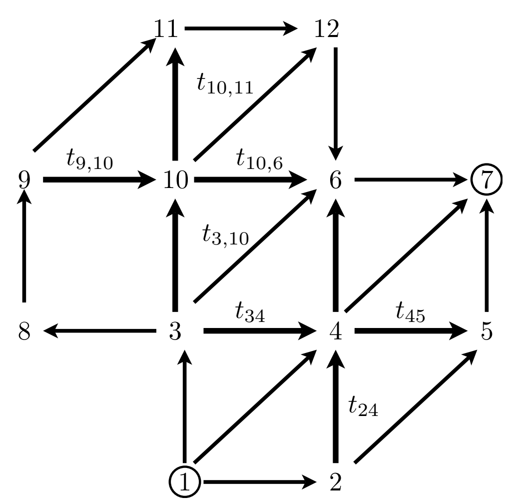


Figure 2.7: Network with multiple cells.

2.7 Summary and future work

In this paper we have analyzed properties of the logit based toll setting problem, and derived sufficient conditions allowing to set tolls to zero or to identical values. This, together with the introduction of appropriate notions of equivalence between networks, leads to simplification rules that yield a reduction of the problem's dimensionality. In particular, a class of unimodal instances has been identified. Furthermore, the notion of network cell allows to identify "parallel tolls", "bottleneck tolls"and "toll cuts"- features of a network's topology leading to simplification rules - by focusing on a subset of the initial graph, thus yielding a richer class of network simplification rules. We conjecture that logit pricing of series-parallel networks is unimodal, and our analysis constitutes a first step in that direction.

As we showed in Section 2.4, toll setting and product line pricing share many theoretical features. Our result according to which parallel toll arcs should be set identical tolls (Theorem 2.4), translates for the product line pricing problem as the known result according to which product prices should be set to generate equal profit margins, for which we provide an original and much shorter proof. From a numerical standpoint however, while approaches developed for product line problems can often be adapted to the toll setting problem, they fail to address arbitrary network structures. For example, the path following algorithm proposed by Hanson in [42] is a local method that requires a well behaved objective function in order for the algorithm to yield a good solution in a global sense. Yet the pricing problem is known to be highly nonconcave, on account of the network constraints. Similarly, the fractional programming approach proposed by Chen [19] looses its key integrality property when network constraints get involved.

In a companion paper, we will develop quasi-optimal numerical methods aimed at solving multicommodity logit pricing problems, by taking advantage of the similarities between stochastic and deterministic formulations of the network pricing problem.

2.8 Notation

$\mathscr{G} = (\mathscr{A}, \mathscr{N})$	network graph
$(o', d', \mathscr{A}')$	network cell associated with the toll arcs $\mathscr{A}'$ with entry and exit nodes o' and d' respectively
$\mathscr{A}_{\text{toll}}$	set of toll arcs (we use $A_{\text{toll}} = \lvert\mathscr{A}_{\text{toll}}\rvert$)
(o, d)	origin - destination couple
$\mathscr{R}$	set of paths from o to d ($R = \lvert\mathscr{R}\rvert$)
$\mathscr{R}_{\text{toll}}$	set of toll paths from o to d ($R_{\text{toll}} = \lvert\mathscr{R}_{\text{toll}}\rvert$)
$\mathscr{R}^{ij}$	set of path segments from node i to node j
$\mathscr{A}(r)$	set of arcs in path r
$\mathscr{A}_{\text{toll}}(r)$	set of toll arcs in path r
$\mathscr{R}(a)$	set of paths incident to arc a
c	vector of fixed costs along the arcs
t	vector of arc tolls
$\tilde{u}_r(t)$	disutility of path r under toll policy t
$u_r(t)$	expectation of $\tilde{u}_r(t)$
$\tilde{\tau}_{ij}^{\mathscr{R}}(t)$	expected disutility of a path segment from node i to node j, given a path set $\mathscr{R}$
$\tau_{ij}^{\mathscr{R}}(t)$	expectation of $\tilde{\tau}_{ij}^{\mathscr{R}}(t)$
$P_r^{\mathscr{R}}(t)$	probability of choosing path $r \in \mathscr{R}$ under toll policy t
$\overline{P}(t)$	transition martix of the Markov chain associated with the network assignment, under toll policy t
$p_a^{\mathscr{R}}(t)$	probability of choosing arc $a \in \mathscr{A}_{\text{toll}}$ under toll policy t, given a path set $\mathscr{R}$
$p_{a\lvert b}^{\mathscr{R}}(t)$	conditional probability of choosing toll arc a, given arc b is chosen, under toll policy t, given a path set $\mathscr{R}$

δ	path-arc incidence matrix
δ_{toll}	toll path - toll arc incidence matrix
θ	logit scale parameter
$\tilde{\varepsilon}_r$	Gumbel error term associated with the disutility of path r
$\mathscr{N}_i^+$	set of successors of node i
$COV(t)$	toll arc choice variance-covariance matrix
$\mathrm{E}_b^a(t)$	elasticity of arc a's choice probability with respect to the toll t_b under toll policy t
$F(t)$	expected toll raised from the commuters
$F^{\mathscr{A}_1}(t\|\mathscr{A}_2)$	expected toll raised from the commuters on the subset of toll arcs $\mathscr{A}_1$ given that the paths chosen are incident to the subset of toll arcs $\mathscr{A}_2$

CHAPITRE 3

A NUMERICAL STUDY OF THE LOGIT NETWORK PRICING PROBLEM

Dans le deuxième article, nous considérons le problème de tarification dans un réseau sous l'affectation logit d'une demande répartie entre plusieurs paires origine-destination. La présence de plusieurs paires origine-destination introduit une combinatoire qui n'était pas présente dans le premier article. Le problème est d'abord abordé dans une perspective numérique et nous développons des outils de résolution efficaces qui permettent d'identifier des solutions quasi-optimales. Bien qu'il soit difficile de garantir l'optimalité globale des solutions trouvées, nous démontrons, sur la base d'arguments numériques que des outils algorithmiques simples permettent de trouver les meilleures solutions. Par ailleurs, le modèle de tarification déterministe joue un rôle important dans l'élaboration de nos algorithmes. Nous caractérisons ensuite, dans un cadre théorique, la parenté relative des modèles de tarification logit et déterministe.

Différentes approximations combinatoires du problème sont à la base des algorithmes que nous avons mis en oeuvre et qui sont définis, alternativement, de sorte à procurer une grande précision localement (autour d'un vecteur de taxes donné), ou globalement sur l'ensemble du domaine de la fonction de revenu. Dans le premier cas, la résolution de l'approximation s'inscrit dans le cadre d'une approche itérative d'optimisation par régions de confiance. Les approximations combinatoires sont résolues de façon exacte lorsque la taille des problèmes n'est pas trop grande ; une heuristique primale-duale est utilisée dans le cas contraire. Dans tous les cas, un algorithme de montée basé sur l'information de premier ordre vient compléter la recherche (méthode de point intérieur).

Les programmes combinatoires ont cet avantage sur les formulations plus générales non linéaires de pouvoir être résolus globalement par énumération. Une question importante que nous soulevons dans le cadre de ce travail est de savoir dans quelle mesure une approximation combinatoire précise est nécessaire afin d'obtenir les meilleures solutions, étant donné un temps et une puissance de calcul limités.

L'approximation combinatoire la plus simple correspond au modèle de tarification déterministe (Programme 1.1). Ce modèle est facile à résoudre (pour les tailles d'instances qui nous intéressent) mais procure une très mauvaise approximation des flots et du revenu logit.

Dans cette optique, nous formulons des programmes combinatoires qui généralisent le modèle de tarification déterministe et permettent théoriquement d'obtenir une approximation du revenu logit d'une précision arbitraire. Ces modèles sont obtenus après avoir linéarisé le système de complémentarité associé aux conditions d'optimalité du niveau inférieur de la formulation bi-niveau exacte (Programme 1.3). Par référence à un ensemble de route $\mathscr{R}$, soit x une affectation des flots telle que $x \in X = \{x \geq 0 | \sum_{r \in \mathscr{R}} x_r = 1\}$ et π le multiplicateur associé. Alors, en utilisant l'opérateur $a \perp b$ pour représenter la complémentarité de deux vecteurs, c.-à-d. $a \perp b \Leftrightarrow a_i b_i = 0 \, \forall i$, nous avons :

$$0 \leq c + t + \frac{1}{\theta} \log x - 1_{\mathscr{R}} \pi \;\perp\; x \in X, \pi \in \mathbb{R}$$

où c est le vecteur des coûts fixes sur les routes, t est le vecteur des tarifs sur les routes, $1_{\mathscr{R}} = [1, \ldots, 1]^T \in \mathbb{R}^{|\mathscr{R}|}$ et où, pour simplifier la notation, une seule paire origine-destination est définie. Deux types d'approximations sont envisagées. La première remplace le logarithme dans le système de complémentarité ci-dessus par une fonction étagée ; la seconde, par une fonction linéaire par morceaux. Nous démontrons, sous des hypothèses faibles (que les segments linéaires ou constants soient de même longueur), que l'erreur d'approximation sur le revenu logit est, dans le premier cas, inversement proportionnelle au nombre de segments constants, et, dans le second cas, inversement proportionnelle au carré du nombre de segments linéaires.

L'algorithme de régions de confiance s'inspire de celui présenté par Marcotte et al. [57] pour la résolution de programmes bi-niveau non linéaires et qui a notamment été implanté avec beaucoup de succès pour la résolution du modèle de tarification déterministe par Colson et al. [26]. Les modèles combinatoires utilisés par Colson et al. se basent sur des approximations de premier ordre des contraintes et de l'objectif bilinéaire. Dans notre cas, une approximation de premier ordre des contraintes permet d'exprimer l'objectif de façon exacte par une forme quadratique définie négative. Nous comparons aussi les résultats obtenus en utilisant une fonction étagée pour linéariser les contraintes, les différents segments constants étant alors définis de façon adaptative à chaque itération. Notre formulation du problème nous permet par ailleurs d'obtenir une expression exacte de l'objectif de niveau supérieur correspondant à une forme quadratique définie négative. Dans tous les cas, il s'est avéré important de prendre des mesures pour éviter, le cas échéant, que les flots itérés ne s'approchent pas trop de zéro, le domaine du logarithme étant relativement ouvert autour de l'origine.

De nombreuses combinaisons algorithmiques se présentent naturellement. Par exemple, la résolution exacte ou heuristique d'une approximation combinatoire donnée, dont la solution vient démarrer une méthode locale basée sur l'information de premier ordre, ou alternativement, l'algorithme de région de confiance bi-niveau défini autour d'un autre modèle combinatoire. Les performances d'un certain nombre de ces combinaisons ont été comparées lors d'essais numériques. Une méthode locale randomisée de type point intérieur à été utilisée à titre de *benchmark*.

Nous avons d'abord effectué des tests numériques pour une gamme de valeurs du paramètre d'échelle logit θ, sur des réseaux de haute densité générés avec des valeurs aléatoires uniformes des paramètres tels que les coûts fixes et les demandes entre les couples origine-destination. Pour une petite valeur de θ, le problème d'optimisation est bien conditionné et une méthode locale est efficace. Pour un grande valeur de θ, les formulations probabiliste et déterministe sont suffisamment apparentées pour que l'optimum de l'approximation combinatoire la plus simple soit dans la bonne région de concavité. Le cas intermédiaire ne s'est pas présenté et les meilleures solutions trouvées s'accordent avec les résultats d'une recherche locale randomisée. Notons que nous nous sommes bornés à considérer des valeurs « raisonnables » de θ. C'est-à-dire des valeurs telles que le revenu logit et l'information de premier ordre puissent être évalués sans entraîner de difficultés numériques.

Ces résultats nous ont encouragés à construire des réseaux spécifiquement pour que le problème soit plus difficile. Un de nos résultats théoriques est d'ailleurs l'existence de réseaux pour lesquels la solution du modèle de tarification déterministe se trouve dans une région de concavité *arbitrairement mauvaise* du modèle de tarification logit. De tels réseaux ont été construits en jumelant des demandes fortes sur des paires origine-destination où les routes de la compétition sont attrayantes avec des demandes faibles sur des paires origine-destination où les routes de la compétition ne sont pas attrayantes. Bien que l'approximation déterministe ne permette pas de trouver de bonnes solutions sur ces réseaux, nos essais numériques montrent, d'une part, qu'une recherche locale randomisée trouve encore de bonnes solutions, et d'autre part, qu'il n'est pas nécessaire de produire une approximation combinatoire très précise du revenu logit pour identifier les meilleures régions de concavité, quatre ou cinq segments constants ou linéaires sont en général suffisants.

Nous avons obtenu des solutions d'une qualité variable avec l'approche par régions de confiance. Comparativement à une méthode n'exploitant que l'information de premier ordre (méthode de point intérieur), la suite des itérés témoigne effectivement d'une moins grande propension à converger vers une solution sous-optimale (au sens global). Cependant, la résolution globale d'une approximation combinatoire suffisamment précise, et qui n'est pas définie localement, permet généralement d'identifier de meilleures régions de concavité. Par ailleurs, de bons résultats ont été obtenus sur les instances faisant intervenir une grande valeur du paramètre d'échelle logit (variance faible) en utilisant l'algorithme de régions de confiance comme méthode locale pour compléter les solutions trouvées par d'autres algorithmes.

Notre analyse permet de tirer un certain nombre de conclusions sur la nature de notre problème qui sont en partie d'ordre qualitatif. D'abord, l'affectation logit des flots induit un effet de lissage sur la fonction du revenu qui favorise les approches locales. Ensuite, la bonne performance d'une recherche locale randomisée montre que pour toute valeur raisonnable du paramètre d'échelle logit, la fonction objectif admet beaucoup moins de maxima que dans le cas déterministe. Finalement, les modèles de tarification logit et déterministe s'accordent souvent sur la région des tarifs optimaux et ne sont donc pas si différents dans leur structure.

Submitted to INFORMS Journal on Computing on January 17, 2012.

NUMERICAL APPROACHES TO THE LOGIT NETWORK PRICING PROBLEM

FRANÇOIS GILBERT, PATRICE MARCOTTE AND GILLES SAVARD

ABSTRACT

In this paper, we address the numerical solution of a pricing problem where users are assigned according to a logit model onto the paths of a transportation network. Although this highly nonconvex problem admits a large number of local optima, we develop strategies that allow to uncover near-optimal solutions, through a mixture of approximations and local ascent techniques.

keywords: Network pricing, discrete choice models, bilevel programming, combinatorial optimization, nonlinear optimization, trust-region.

January 2012

3.1 Introduction

This work is concerned with a profit maximization model that takes place over a transportation network, and fits the pricing literature [12, 16, 20, 22, 36, 47, 76, 81]. While the problem under consideration has been investigated within a deterministic framework by Marcotte et al. in [51], little attention has been paid to its stochastic variant, where users are assigned to paths of the network according to a discrete choice model, although a notable exception can be found in the product line design and congestion pricing literature [42, 82].

In this paper, the focus is on the development of efficient numerical procedures for obtaining near-optimal solution to the logit-constrained pricing problem. Since the deterministic version has been proved to be NP-hard, it follows that the stochastic extension is NP-hard as well (see Roch et al. [65]). However, the combinatorial nature of the former makes it amenable to mixed integer formulations that can be solved via enumeration approaches, such as branch-and-bound or branch-and-cut. Our aim is to exploit the similarities between the stochastic and deterministic formulations to devise efficient algorithms for network logit pricing. For instance, both models are readily compared within a bilevel environment, which is natural under the deterministic formulation, but can also be achieved in the stochastic case by expressing the logit assignment as the solution of an entropy maximization problem. This representation suggests simple and efficient combinatorial approximations to the logit revenue function.

Within this framework, two complementary resolution schemes are considered. The first scheme consists in warm starting a local search method from the global optimum of some combinatorial approximation. This can be achieved in several ways that will be explored throughout the paper. The second scheme is a bilevel trust-region method based on local combinatorial approximation. Actually, the two approaches can mingle to initiate the trust-region algorithm from the solution provided by the first scheme.

In parallel with algorithmic development, we prove theoretical error bounds for the various heuristic methods considered, with respect to the accuracy of the approximations. We also assess the effect of the smoothing induced by the random term on the convergence of the various methods to a good solution, in the global optimization sense.

The rest of the paper is organized as follows. In Section 3.2, we provide a brief overview of the deterministic network pricing problem, which is set within the framework of bilevel programming or, equivalently, Stackelberg games. In Section 3.3, we introduce network logit pricing, and make a parallel with its deterministic counterpart. In Section 3.4, we discuss a heuristic procedure based on a deterministic approximation of the logit model and contrast its performance against that of a local search initiated at the origin. In Section 3.5 we consider the implementation of a bilevel trust-region algorithm constructed around a combinatorial model. The next two sections (Sections 3.6–3.7) are devoted to approximation schemes based on a piecewise linear approximation of the logarithmic function, either discontinuous (step function) or continuous. We show that this approach yields asymptotically optimal solutions, and numerical results support the claim that a coarse grain discretization is sufficient for uncovering quasi-optimal solutions. We also provide a rigorous proof that, under a reasonable set of assumptions, and as the number of "pieces" increases, the solution converges to the optimal solution of the original problem. We also provide an estimate of the error bound with respect to the number of pieces. In Section 3.8, additional numerical results are presented which emphasize the impact of both the logit parameter (inversely related to the variance) and the topology of the network on the performance of the various algorithms. Indeed, large variances induce smooth objective functions that lend themselves naturally to local ascent methods, while small variances reveal the combinatorial nature of the problem and make it closer to its deterministic counterpart. Closing remarks are given in Section 3.9. Notations are summarized in Section 3.10.

3.2 Deterministic toll setting

Let $\mathscr{G} = (\mathscr{N}, \mathscr{A})$ be a directed graph, $\mathscr{A}_{\text{toll}} \subset \mathscr{A}$ a subset of toll arcs, $\mathscr{Q} \subset \mathscr{N} \times \mathscr{N}$ a set of origin-destination pairs, and $\{\mathscr{R}^q\}_{q \in \mathscr{Q}}$ the sets of paths, one set per origin-destination pair. Throughout the paper, several graphs and sets are denoted in calligraphic fonts and, in the latter case, we let $A = |\mathscr{A}|$, generically.

A constant demand of d_q commuters is associated with origin-destination $q \in \mathscr{Q}$, and is assigned to paths $\mathscr{R}^q$. To each arc $a \in \mathscr{A}$ is assigned a fixed cost c_a^A and, if $a \in \mathscr{A}_{\text{toll}}$, a toll t_a^A. This structure induces, along a path r relevant to an origin-destination pair q of the network, the quantities $c_r = \sum_{a \in r} c_a^A$ and $t_r = \sum_{a \in r} t_a^A$, where the sums are taken over the arcs incident to path $r \in \mathscr{R} = \cup_{q \in \mathscr{Q}} \mathscr{R}^q$. Also let $c^q, t^q \in \mathbb{R}^{R^q}$ represent the restrictions of c and t to the set of indices in $\mathscr{R}^q$ for each $q \in \mathscr{Q}$. The notation $q(r) \in \mathscr{Q}$ refers to the origin–destination pair to which path $r \in \mathscr{R}$ belongs. In the sequel, $1_{\mathscr{Q}}$ and $1_{\mathscr{R}}$ are vectors of ones of dimension Q and R respectively, and matrix $1_{\mathscr{R}}^{\mathscr{Q}} \in \{0,1\}^R \times \{0,1\}^Q$ is such that for $z \in \mathbb{R}^R$ and $y \in \mathbb{R}^Q$ we have $z = 1_{\mathscr{R}}^{\mathscr{Q}} y \Leftrightarrow z_r = y_{q(r)}$, $r \in \mathscr{R}$.

The network pricing problem consists in determining tolls on the subset of arcs $\mathscr{A}_{\text{toll}}$ so as to maximize the revenue generated by the demand, i.e., the expected toll revenue raised from commuters. In the deterministic framework, flows are concentrated on shortest origin-destination paths. This yields a bilinear optimization problem with complementarity constraints which can be reformulated as a mixed integer linear program in either arc or path flow space.

The toll problem is best expressed as a bilevel program. Let X be the set of unit demand assignments over all OD pairs:

$$X = \{x \in [0,1]^{R^q} | \sum_{r \in \mathscr{R}^q} x_r = 1,\, q \in \mathscr{Q}\}, \tag{3.1}$$

and $T \subset \mathbb{R}^R$ the set of bounded feasible path tolls of diameter $\operatorname{diam} T < \infty$. We have $t \in T \Leftrightarrow \exists t^A \in \mathbb{R}^{A_{\text{toll}}}$ such that

$$t_r = \sum_{a \in \mathscr{A}_{\text{toll}}(r)} t_a^A.$$

The sets T and X are the upper and lower level feasible sets, respectively, of the following bilevel formulation of the deterministic toll setting problem:

Program 3.1.

$$\max_{t,x} \quad f(t,x) = \sum_{r\in\mathscr{R}} d_{q(r)}\, t_r\, x_r \tag{3.2}$$

$$\text{s.t.} \quad t \in T,\ x \in X$$

$$x \in \arg\min_{y}\{g^{\text{det}}(t,y) = (c+t)\cdot y \mid y \in X\}.$$

where the upper level objective f is the revenue and lower level objective g^{det} is the sum of the disutilities of the paths chosen on each origin-destination pairs. According to this formulation, assignment x is an optimal solution of the lower level optimization program, and is solved by determining shortest paths, one per OD pair. This framework fits the definition of a Stackelberg game involving a revenue-maximizing leader and a cost-minimizing follower, and where the leader anticipates the follower's response x to his own strategy t. Program 3.1 has been introduced by Labbé et al. [51] and solved to global optimality by implicit enumeration techniques. As mentioned by Labbé et al., the lower level optimal solution is in general degenerate. Indeed, suppose otherwise and let (t^*,x^*) be optimal and assume that no two paths have equal disutilities on any origin-destination pair. We can assume that at least one $r \in \mathscr{R}$ is such that $t^*_r > 0$ and $x^*_r \neq 0$, which implies that $\mathscr{R}^{q(r)} \ni r' \neq r \Rightarrow x_{r'} = 0$. Consider the toll vector t' such that $r' \neq r \Rightarrow t'_{r'} = t^*_{r'}$ and $t'_r = t^*_r + \delta$ for some $\delta > 0$ sufficiently small for the assignment to remain unchanged, that is $\|t^* - t'\| < \delta \Rightarrow x^* \in \arg\min\{g^{\text{det}}(t',y)|y \in X\}$. But then we have $f(t^*,x^*) < f(t',x^*)$, a contradiction. Thus at least two paths have equal disutility at the optimum, and maximization of f at the upper level insures that the prevailing one is the one that best suits the leader.

This formulation is in line with the so-called optimistic interpretation of the bilevel program, where the lower level solution is the one that yields the highest revenue for the leader. In contrast, the pessimistic interpretation assumes that the lower level response should be the one that least suits the leader. We observe that Program 3.1 is compatible with both interpretations. Indeed, as observed by Labbé et al. [51], an appropriate perturbation of a local optimum yields a non-degenerate lower level response and, in turn, an upper level objective value as close as desired from the optimistic optimum. The result is proved for the sake of completeness.

Theorem 3.1. *Let (t,x) denote an optimal solution of the deterministic network pricing problem (Program 3.1). Then there exists a toll policy t' such that all shortest paths with respect to $c+t'$ have identical tolls (for any given OD pair), and*

$$|f(t,x)-f(t,x')| \leq \varepsilon, \; \forall x' \in \arg\min_{y\in X}\{g^{\det}(t',y)\}.$$

Proof. Let t^A be such that $t_r = \sum_{a\in r} t^A_a, r \in \mathscr{R}$ with $x^A = \sum_{r:a\in r} x_r$, and select $t' \in T$ such that $t'^A_a = t^A_a/(1+\delta)$ for all $a \in \mathscr{A}_{\text{toll}}$, where $t'_r = \sum_{a\in r} t'^A_a$ whenever $a \in \mathscr{A}$, $r \in \mathscr{R}$ and $\delta > 0$. The perturbation is such that the paths with the highest tolls are the most penalized. Then, for any positive δ, all shortest paths under t' have the same toll, which can be made arbitrarily close to the toll on the most profitable shortest path under t'. □

As an immediate corollary, we obtain that the supremum of the pessimistic objective (which might not be achievable) is equal to the optimum of the optimistic objective.

Upon replacing the lower level of Program 3.1 by its first-order optimality conditions, we readily obtain a single level mathematical program involving linear complementarity constraints. As the lower level objective $g^{\det}$ is separable by OD pair, and since strong duality prevails, its objective can be expressed as

$$\begin{aligned} g^{\det}(t,x) &= (c+t)\cdot x \\ &= 1_{\mathscr{Q}}\cdot\pi, \end{aligned} \tag{3.3}$$

where the components of vector π are the multipliers associated with the equality constraints in the definition of X in equation (3.1). It is well known that π_q is equal to the disutility of the shortest path from o_q to d_q. Using (3.3) to linearize the upper level bilinear objective (3.2) we obtain the mixed integer program

Program 3.2.

$$\begin{aligned} \max_{t,x,\pi} \quad & \sum_{r\in\mathscr{R}} d_{q(r)}[\pi_{q(r)} - c_r x_r] \\ \text{s.t.} \quad & 0 \leq c + t - 1_{\mathscr{Q}}^{\mathscr{R}}\pi \perp x \geq 0 \qquad (3.4) \\ & t \in T, x \in X. \end{aligned}$$

Upon introduction of a "big-M" constant and the binary vector $z \in \{0,1\}^R$, one derives the equivalent MIP formulation

Program 3.3.

$$\begin{aligned} \max_{t,x,\pi} \quad & \sum_{r\in\mathscr{R}} d_{q(r)}[\pi_{q(r)} - c_r x_r] \\ \text{s.t.} \quad & 0 \leq c + t - 1_{\mathscr{Q}}^{\mathscr{R}}\pi \leq M 1_{\mathscr{Q}}^{\mathscr{R}} z \qquad (3.5) \\ & x \leq 1_{\mathscr{R}} - z \\ & t \in T, x \in X \end{aligned}$$

that can be solved for its global solution by a commercial software such as CPLEX [1], the computational effort depending on the tightness of M. The larger the M, the worst the quality of the associated linear relaxation and, in turn, the CPU time required to prune the implicit enumeration tree. A suitable value of M corresponds to the largest value that can take any given component of the vector $c + t - 1_{\mathscr{Q}}^{\mathscr{R}}\pi$ as t runs over T and π is a feasible dual vector. As π_q takes the values of the disutility of the shortest path between o_q and d_q, we should have: $M \geq u_r(t) - u_\ell(t), r, \ell \in \mathscr{R}^q, q \in \mathscr{Q}, t \in T$. Alternatively, one can introduce origin-destination specific values for M to get a tighter linear relaxation of Program 3.3. For further information concerning good choices of M, the reader is referred to Dewez et al. [33] or Heilporn et al. [43]. For simplicity, we will assume in the sequel a common value M for all OD pairs.

3.3 Logit toll setting

The logit network pricing problem is a smooth perturbation of the deterministic model where the "all or nothing" assignment is replaced by a probabilistic assignment over all paths of the network. In the logit framework, commuters minimize their disutility

$$\tilde{u}_r = c_r + t_r + \tilde{\varepsilon}_r,$$

where $\tilde{\varepsilon}_r$ are i.i.d. Gumbel variates:

$$\tilde{\varepsilon}_r \sim \text{Gumbel}(-\eta/\theta, \theta).$$

In the above, the scale parameter θ is positive, η denotes Euler's constant and $\mathbb{E}[\tilde{u}_r] = u_r$ with $u_r = c_r + t_r$. Since Gumbel variates belong to the family of extreme value distributions, we have

$$\min_r \{\tilde{u}_r | r \in \mathscr{R}'\} \sim \text{Gumbel}(-\theta^{-1} \log \sum_{r \in \mathscr{R}^q} \exp(-\theta u_r) - \eta/\theta, \theta)$$

for any nonempty subset $\mathscr{R}'$ of $\mathscr{R}$. Given $r \in \mathscr{R}'$, the probability measure $\mathbb{P}(\tilde{u}_r \leq \tilde{u}_\ell | \ell \in \mathscr{R}')$ is expressed as

$$\text{logit}_r(t|\mathscr{R}') = \frac{\exp(-\theta(c_r + t_r))}{\sum_{\ell \in \mathscr{R}'} \exp(-\theta\,(c_\ell + t_\ell))},$$

where the scale parameter θ impacts the relative spread of the logit assignment among available paths. For a given toll vector, a null value yields an assignment where flow is evenly spread over all paths. A larger value yields an assignment where flow is concentrated on paths of smallest disutility. In the limit $\theta = \infty$, flow is restricted to shortest paths. A detailed account of the logit route choice model is found in Ben-Akiva and Bierlaire [9].

The logit toll setting problem takes the form of the unconstrained mathematical program

Program 3.4

$$\max_{t \in T} \quad f^{\text{logit}}(t) = \sum_{q \in \mathscr{Q}} d_q \sum_{r \in \mathscr{R}^q} \text{logit}_r(u|\mathscr{R}^q)\, t_r. \tag{3.6}$$

The following closed form expression of the gradient, in terms of the arc-toll vector $t^A \in \mathbb{R}^{A_{\text{toll}}}$, is given without proof:

$$\nabla f^{\text{logit}}(t^A) \quad = \quad \sum_{q \in \mathscr{Q}} d_q \left[p(t|\mathscr{R}^q) - \theta\, \text{COV}(t^A|\mathscr{R}^q)\, t^A \right], \tag{3.7}$$

where $p \in \mathbb{R}^{A_{\text{toll}}}$ is the toll arc choice probability vector induced by the logit path choice probabilities, and $\text{COV} \in \mathbb{R}^{A_{\text{toll}}} \times \mathbb{R}^{A_{\text{toll}}}$ denotes the associated variance-covariance matrix. The reader is referred to Gilbert et al. [39] for a detailed sensitivity analysis of the logit network pricing problem.

Alternatively, the logit distribution can be characterized as the solution of a convex optimization problem. This allows to recast Program 3.4 within a bilevel programming framework, more readily compatible with the deterministic formulation:

Program 3.5.

$$\begin{aligned} \max_{t,x} \quad & f(t,x) \\ \text{s.t.} \quad & t \in T \\ & x \in \quad \arg\min_{y \in X} \{ g(t,y) = (c+t) \cdot y \; + \frac{1}{\theta} y \cdot \log y \mid y \in X \}, \end{aligned}$$

where the logarithm is taken component-wise and, by continuity, we set $0 \log 0 = 0$.

Let $L(t,x,\pi)$ be Program 3.5's lower level Lagrangian function, where π is the vector of multipliers associated with the simplex constraints $y \in X$. We have

$$\begin{aligned} \nabla_x L(t,x,\pi)^T &= \nabla_x g(t,x|\theta)^T - 1_{\mathscr{R}}^{\mathscr{Q}} \pi \\ &= \frac{1}{\theta}(\log x + 1_{\mathscr{R}}) + c + t - 1_{\mathscr{R}}^{\mathscr{Q}} \pi, \end{aligned}$$

where the exponential is also taken component-wise and which yields

$$\nabla_x L(t,x,\pi)^T = 0 \Leftrightarrow x = \exp[-\theta(c + t - 1_{\mathscr{R}}^{\mathscr{Q}} \pi)].$$

Substituting this last expression for x in the simplex constraints yields

$$\pi_q = -\theta^{-1} \log \sum_{r \in \mathscr{R}^q} \exp[-\theta(c_r + t_r)],$$

from which the equivalence of Program 3.4 and Program 3.5 follows.

Note that the lower level relates to entropy maximization, which is at the core of an information theoretic derivation of the logit choice model (see Erlander [37]). Actually Fisk [38], alongside many others, expressed the logit assignment in a congested network as the solution of an entropy maximization problem [3, 55, 69]. Note also that the lower level gradient of Program 3.5 is unbounded at the origin, and thus the term $x \log x$ acts as a barrier function, forcing path probabilities away from zero. This feature allows to introduce a constraint that cuts away the origin, but not the optimal solution. To this aim, let $\operatorname{diam} c$ denote the largest difference between path fixed costs:

$$\operatorname{diam} c = \max\{c_r | r \in \mathscr{R}\} - \min\{c_{r'} | r' \in \mathscr{R}\}. \tag{3.8}$$

With reference to any feasible solution (t,x) of Program 3.5 let

$$\delta_{\min} = \exp[-\theta(\operatorname{diam} c + \operatorname{diam} T)]/R. \tag{3.9}$$

Then for any $r \in \mathscr{R}$ we have

$$\begin{aligned} x_r &= \frac{\exp[-\theta(c_r + t_r)]}{\sum_{r' \in \mathscr{R}} \exp[-\theta(c_{r'} + t_{r'})]} \\ &\geq \frac{\exp[-\theta(\max\{c_{r'+} + t_{r'+} | r'+ \in \mathscr{R}\})]}{R \exp(-\theta(\min\{c_{r''} + t_{r''} | r'' \in \mathscr{R}\})]} \\ &> \delta_{\min}, \end{aligned} \tag{3.10}$$

which is the desired "cut". For algorithmic purposes, we can then replace X by $X_0 = \{x | x \geq \delta_{\min} 1_{\mathscr{R}}\}$ in Program 3.5. We will also assume that the set T is bounded, which is the case for all practical purposes.

The bilevel formulation of Program 3.5 differs from its deterministic counterpart (Program 3.1) only in the presence of the entropic term located at the lower level of Program 3.5. It is then natural to develop an algorithmic framework based on replacing the entropic term by an expression that is more suitable to global optimization, and the following variational formulation will prove useful in that respect. With reference to Program 3.5's lower level objective g, set

$$F(t,x) = \nabla_x g(t,x) \tag{3.11}$$

and let $\mathrm{SOL}(F(t,\cdot),X)$ be the solution set associated with the variational inequality $\mathrm{VI}(F(t,\cdot),X)$, that is, $\mathrm{SOL}(F(t,\cdot),X) = \{y \in X | F(t,y) \cdot (x-y) \geq 0, \forall x \in X\}$. The logit network pricing problem can then be expressed as

Program 3.6.

$$\begin{aligned} \max_{t,x} \quad & f(t,x) \\ \text{s.t.} \quad & t \in T \\ & x \in \mathrm{SOL}(F(t,\cdot),X). \end{aligned} \tag{3.12}$$

Using (3.10) we have $x \in \mathrm{SOL}(F(t,\cdot),X) \Rightarrow x \in \mathrm{SOL}(F(t,\cdot),X_0)$. Then

$$F(t,x) = \nabla_x g(t,x) = c + t + \theta^{-1}(\log x + 1_{\mathscr{R}})$$

and so the associated Jacobian matrix is uniformly positive definite over X_0 with largest eigenvalue $(\theta\delta_{\min})^{-1}$ and smallest eigenvalue θ^{-1}. The following result will be useful in the sequel.

Lemma 3.1. *Let (t,x) be feasible for Program 3.6 and $\bar{x} \in \mathrm{SOL}(\overline{F}(t,\cdot),X_0)$ where $\overline{F}$ is an approximation of F. Set $d^{\max} = \max\{d_q \mid q \in \mathscr{Q}\}$, then*

$$|f(t,x) - f(t,\bar{x})| \le \|F(t,\bar{x}) - \overline{F}(t,\bar{x})\|\, d^{\max}\, \mathrm{diam} T/\theta.$$

Proof. The result is shown for the unit demand and single origin-destination case, generalization to the case with multiple OD pairs being straightforward. As X_0 is compact and F is strongly monotonic with modulus θ, we have that

$$\|x - \bar{x}\| \le \theta^{-1}\|F(t,\bar{x}) - \overline{F}(t,\bar{x})\|.$$

Then

$$\begin{aligned}
|f(t,\bar{x}) - f(t,x)| &= \sum_{q\mathscr{Q}} d_q t^q \cdot (\bar{x}^q - x^q) \\
&\le d^{\max}\|t\|\|\bar{x} - x\| \\
&\le d^{\max}\mathrm{diam}\ T\ \|F(t,\bar{x}) - \overline{F}(t,\bar{x})\|/\theta.
\end{aligned}$$

□

Our final reformulation involves a single equality constraint for each OD pair, and will also lend itself to approximations that can be solved for their global optima:

Program 3.7.

$$\max_{x,t,\pi} \quad \sum_{q\in\mathscr{Q}} d_q \left(-c^q \cdot x^q - \frac{1}{\theta} x^q \cdot \log x^q + \pi_q \right) \tag{3.13}$$

$$\text{s.t.} \quad c + 1^{\mathscr{R}}_{\mathscr{R}_{toll}} t + \frac{1}{\theta} \log x - 1^{\mathscr{Q}}_{\mathscr{R}} \pi = 0 \tag{3.14}$$

$$x \in X, \, t \in T.$$

Its validity rests on (3.14). Indeed, let $x = \exp[-\theta(c+t-1^{\mathscr{Q}}_{\mathscr{R}}\pi)]$, where the components of π act as normalization factors that insures that $x^q \in \text{logit}(t|\mathscr{R}^q)$ for each $q \in \mathscr{Q}$. Multiplying both sides of (3.14) by x yields $t \cdot x = -(c+t)\cdot x - \frac{1}{\theta} x \cdot \log x + 1^{\mathscr{Q}}_{\mathscr{R}}\pi$, from which it follows that $f(t,x) = \sum_{q\in\mathscr{Q}} d_q(-c^q \cdot x^q - \frac{1}{\theta} x^q \cdot \log x^q + \pi_q)$, where f is Program 3.5's upper level objective (the logit revenue function). The equivalence with Program 3.7 follows.

Note that constraint (3.14) is equivalently expressed as:

$$0 \le c + t + \frac{1}{\theta} \log x - 1^{\mathscr{Q}}_{\mathscr{R}} \pi \perp x \in X, \tag{3.15}$$

where complementarity is trivially satisfied since $\log x > -\infty \Rightarrow x > 0$. Dropping the logarithm in (3.15) yields the deterministic assignment. We will use the deterministic assignment in Section 3.4 to construct an approximation of the logit toll setting problem. In Section 3.5, the logarithmic term in (3.15) will be replaced by a first-order local approximation, yielding a quadratic concave approximation of the logit revenue function. This approximation will be embedded within a trust region framework. In Sections 3.6 and 3.7 the logarithm will be replaced by a step function and a piecewise linear function, respectively, which in both case lead to a concave piecewise linear approximation of the logit revenue.

Note that the entropic barrier function induced by the logit assignment provides a smoothed version of its deterministic counterpart. It is closely related to the logarithmic barrier used by Dussault et al. in [36] for solving the deterministic problem through a path following approach, the difference being that the "barrier" parameter is an integral part of the network pricing problem, and is consequently not an *ad hoc* parameter introduced for algorithmic purposes.

We close this section by stating a result that makes the connection between the deterministic and the limiting case of the logit version of the network pricing problem.

Theorem 3.2. *Let f^{det} be the deterministic revenue (objective of Program 3.1) and f_θ^{logit} be the logit revenue. Define $\varphi^*_{\text{logit}}(\theta) := \max_{t\in T} f_\theta^{\text{logit}}(t)$ and $\varphi^*_{\text{det}} := \max_{t\in T} f^{\text{det}}(t)$. If T is compact, then we have that*

$$\lim_{\theta\to\infty} \varphi^*_{\text{logit}}(\theta) = \varphi^*_{\text{det}}.$$

Proof. First, we show that $\lim_{\theta\to\infty} \varphi^*_{\text{logit}}(\theta) \geq \varphi^*_{\text{det}}$. Let t^* be such that $f^{\text{det}}(t^*) = \varphi^*_{\text{det}}$. By Theorem 3.1, for any $\varepsilon > 0$ there exists a perturbation t' of t^* such that (i) $f^{\text{det}}(t') \geq \varphi^*_{\text{det}} - \varepsilon$ and (ii) $r, r' \in \mathscr{R}^q \Rightarrow t_r = t_{r'}, q \in \mathscr{Q}$. As logit probabilities, for fixed t, concentrate on shortest paths for increasing values of θ, and on account of the fact that under t' all shortest paths have identical toll values, we obtain that $\lim_{\theta\to\infty} f^{\text{logit}}(t'|\theta) \geq \varphi^*_{\text{det}} - \varepsilon$, which immediately yields

$$\lim_{\theta\to\infty} \varphi^*_{\text{logit}}(\theta) \geq \varphi^*_{\text{det}} - \varepsilon.$$

Next, let us show that $\lim_{\theta\to\infty} \varphi^*_{\text{logit}}(\theta) \leq \varphi^*_{\text{det}}$, which will yield the desired result. To this aim, let us consider sequences $\theta_k \to \infty$ and $t_k \in \arg\max_t f^{\text{logit}}(t|\theta_k)$ with $x_k := \text{logit}(t_k|\theta_k)$. By compactness of T, there exists a subsequence K of indices and a toll policy $\bar{t} \in T$ such that $\lim_{k\in K} t_k = \bar{t}$. By continuity of the logit probabilities there also exists an assignment $\bar{x}$ such that $\lim_{k\in K} \text{logit}(t_k|\theta_k) = \bar{x}$. For some $q \in \mathscr{Q}$, consider paths $r, r' \in \mathscr{R}^q$ and a positive number δ such that

$$u_r(\bar{t}) \leq u_{r'}(\bar{t}) - \delta.$$

Then there exists an index $\bar{k}$ such that $k > \bar{k}$ yields (i) $\delta + u_r(t_k) \leq u_{r'}(t_k)$ and (ii)

$\lim_{k\geq\bar{k}} \text{logit}_{r'}(t_k,\theta_k) = 0$, where the last equality follows from

$$\begin{aligned}
\text{logit}_{r'}(\bar{t},\theta) &\leq \frac{\exp[-\theta u_{r'}(\bar{t})]}{\exp[-\theta u_{r'}(\bar{t})] + \exp[-\theta u_r(\bar{t})]} \\
&\leq \frac{\exp[-\theta(u_r(\bar{t})+\delta)]}{\exp[-\theta(u_r(\bar{t})+\delta)] + \exp[-\theta u_r(\bar{t})]} \\
&= \frac{\exp(-\theta\delta)}{\exp(-\theta\delta)+1},
\end{aligned}$$

which goes to zero as θ goes to infinity. Since only shortest paths carry positive flow, the solution $(\bar{t},\bar{x})$ is feasible for Program 1 and $\lim_{\theta\to\infty} \varphi^*_{\text{logit}}(\theta) \leq \varphi^*_{\text{det}}$, as desired. □

3.4 A deterministic heuristic

In this section, we present a two-phase optimization procedure. In the first phase the deterministic formulation (Program 3.1) is solved to global optimality. In the second phase, a local search is performed with respect to the true function, and is initiated from the solution obtained in the first phase. To what extent the deterministic formulation provides a good approximation of the logit revenue function, in the sense that solving the former allows to reach a region of the latter from where a global optimum can be reached, is the question that arises. While intuition suggests that this procedure should find good solutions for either very large or very small values of the parameter θ, its performance in general is hard to assess *a priori*, as shown by the following example.

The left-hand side of Figure 3.1 plots the deterministic and logit revenue functions parameterized by $\theta = 0.5$ and $\theta = 2.0$ corresponding to the network and data shown on the right-hand side, and highlights a counter-intuitive result. For $\theta = 2.0$, the logit revenue function is bimodal, with a global optimum on the right while, for $\theta = 0.5$, the function's unique critical point (global optimum) appears on the left. In the former situation, characterized by a small variance, initiating the local search from the solution of the deterministic optimum (located on the left) will fail to identify the right optimum.

The impact of the fixed cost structure on the performance of the heuristic is also

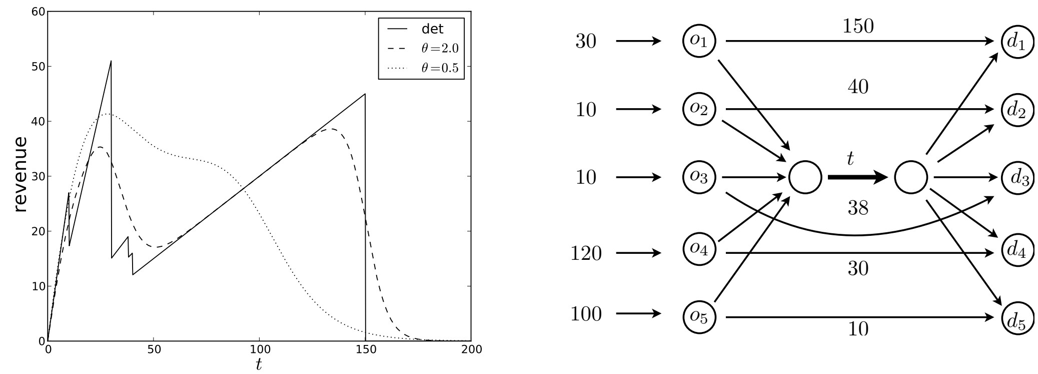


Figure 3.1: Logit (dashed lines) and deterministic (solid line) revenue functions and the associated network.

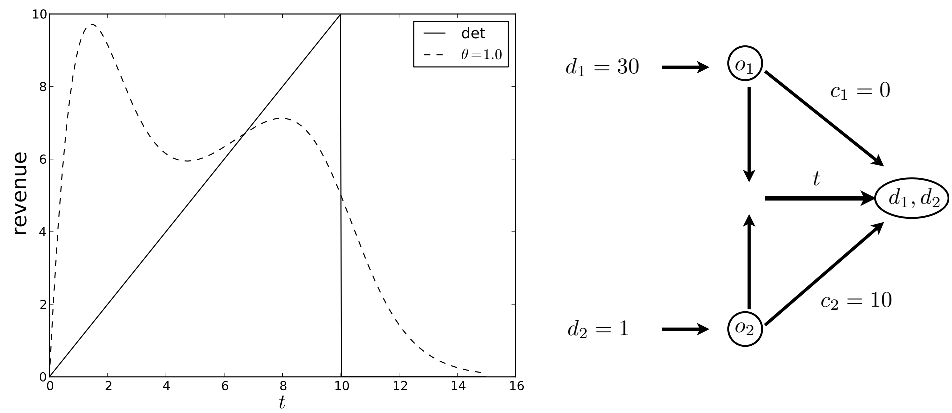


Figure 3.2: Logit (dashed line) and deterministic (solid line) revenue functions and the associated network.

important. More precisely, we observed that large demands associated with small free toll path fixed costs, together with small demands associated with large free toll path fixed costs, frequently results in poor performance of the heuristic. This is illustrated on the network of Figure 3.2 which involves two OD pairs and two paths for each OD pair. The deterministic approximation is blind to the commuters associated with OD pair (o_1, d_1), the competition fixed cost being smaller than that of the toll path under *any* positive toll value. This is not an isolated case: setting c_1 to any value smaller than $10/31$ also misleads the deterministic heuristic (the revenue curve for $c_1 = 1/4$ is given on the right-hand side). These observations lead to the following result.

Theorem 3.3. *There exist instances such that the the solution found by the deterministic heuristic is arbitrarily worse than the optimal value of the original problem.*

Proof. Consider the network of Figure 3.2, and let

$$c_1 < d_2c_2/(d_1+d_2). \tag{3.16}$$

Let $f^{\det(t)}$ be the deterministic revenue associated with toll t and let H a slightly modified Heaviside step function:

$$H(x) \;=\; \begin{cases} 1 & \text{if } x \geq 0 \\ 0 & \text{otherwise.} \end{cases}$$

Then:

$$\begin{aligned} \max_t f^{\det}(t) &= \max_t d_1H(t-c_1)t + d_2H(t-c_2) \\ &= \max\{(d_1+d_2)c_1, d_2c_2\} \\ &= d_2c_2. \end{aligned}$$

Let $f^{\text{logit}}(t)$ denote the logit revenue. We have:

$$\begin{aligned} f^{\text{logit}}(t) &= \frac{d_1\exp(-\theta t)t}{\exp(-\theta t)+\exp(-\theta c_1)} + \frac{d_2\exp(-\theta t)t}{\exp(-\theta t)+\exp(-\theta c_2)} \\ &> \frac{d_1\exp(-\theta t)t}{\exp(-\theta t)+\exp(-\theta c_1)}. \end{aligned}$$

This lower bound on the logit revenue can be made arbitrarily large by way of d_1. Furthermore, for any d_1, bounded values of d_2 and c_2 can be found to satisfy inequality (3.16). The conclusion follows. □

Next, we illustrate a case where the deterministic approximation performs well. Consider an extended version of the network of Figure 3.3 involving 50 OD pairs, two paths per origin-destination pair and a single toll arc. The deterministic and logit revenue curves are shown on Figure 3.3. While it can be argued that increasing the number of OD pairs should yield a more challenging optimization problem, it is also clear, however, that the deterministic approximation can perform much better when the number of OD pairs is increased.

On *Voronoi* or *Delaunay* topologies, that have been used in the pricing literature for their adequation at modelling telecommunication or transportation networks, the deterministic heuristic, when compared to the more elaborate schemes introduced later, fails to find good solutions only on a small number of instances, and performs exceptionally well on the vast majority of them. But then again so does a randomized local search involving only 100 replications. For different values of the scale parameter θ, either the problem is smooth enough for the randomized local search to find the best solution, or it is numerically stiff in the sense that the gradient is very difficult to evaluate. In the later case, the randomized search then more or less reduces to a grid search. In all cases the deterministic heuristic finds the best solutions.

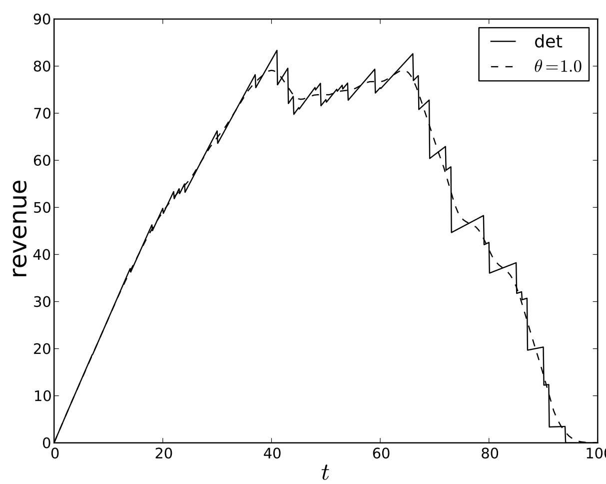


Figure 3.3: Logit (dashed line) and deterministic (solid line) revenue function for a single toll network with 50 origin-destination couples.

However, we did identify a class of instances that are troublesome for the deterministic heuristic. These *circular networks* (see Figure 3.4) are characterized by a high density, each toll path alternating between toll and toll-free arcs. They involve no dominated paths in the sense of Bouthou et al. [12], or network cells in the sense of Gilbert et al. [39]. We generated random circular networks involving either 5 or 10 toll arcs. Five origin-destination pairs with five paths each are defined for the 5-toll instances, and 10 origin-destination pairs with 10 paths each for the 10-toll instances. Paths are pre-selected so as to form a set of minimal fixed costs in which a single toll-free path is present, and all arcs appear at most once on any path. Origins are selected randomly among odd-numbered nodes and destinations among even-numbered nodes.

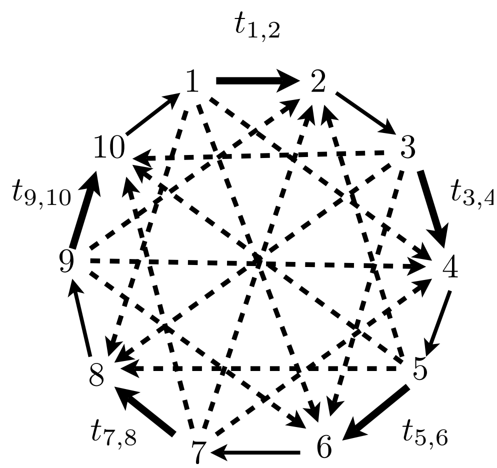


Figure 3.4: Five toll arcs circular network.

Two schemes have been devised to generate fixed costs and demand. Both schemes set to zero all fixed costs on arcs along the diameter, and set demands according to a uniform random variable with support $(1, 100)$. In the first scheme, fixed costs on other toll-free arcs are set according to a uniform variate of equal support diameter, and whose minimal value is proportional to the number of arcs crossed along the diameter. Identifiers associated with instances generated with the first scheme are prefixed with the letter “A” followed by the number of toll arcs.

The second scheme applies the following transformation to the fixed costs and demands generated with the first scheme: each origin-destination demand is either multiplied by 1000 with probability 0.5, or divided by 1000. The associated competition fixed costs are either divided by 1000, in the first case, or multiplied by 1000 in the second case. The rationale behind this strategy is to induce an effect similar to the one observed

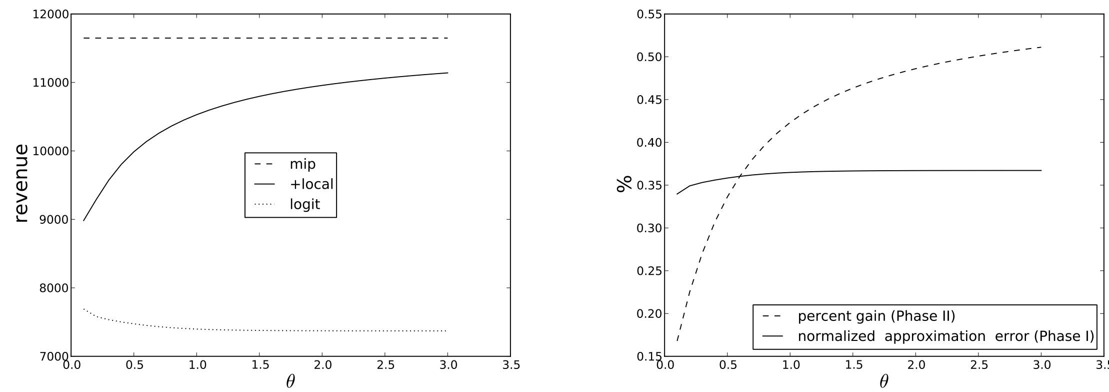


Figure 3.5: Impact of θ on the deterministic heuristic (instance A-10-1): Phase I and Phase II solution values (left); normalized approximation error and percentage gained from Phase II (right).

on the network of Figure 3.2. Identifiers associated with instances generated with the second scheme are prefixed with the letter "B" followed by the number of toll arcs.

Numerical experiments have been conducted on a wide range of values of θ. This is accounted for in Section 3.8. All experiments were conducted on a dual core Pentium D processor running at 3.20 GHz with 4 GB of memory. In Figure 3.5, we assess the impact of θ on the performance of the heuristic on instance A-10-1. On the left-hand side are compared the Phase I and Phase II objectives and the corresponding combinatorial model optimal values. As θ increases, the Phase I solution barely changes, while the value of the Phase II solution and that of the model seem to agree. On the right-hand side, the normalized approximation errors and the percentage gain from Phase II are defined, respectively, as

$$\begin{aligned}
\text{Phase I error} &= \frac{\text{Phase I MIP value} - \text{Phase I logit revenue}}{\text{Phase I MIP value}}; \\
\text{Phase II gain} &= \frac{\text{Phase II logit revenue} - \text{Phase I logit revenue}}{\text{Phase I logit revenue}}.
\end{aligned}$$

The general behaviour shown in Figure 3.5 is typical: the deterministic model provides a poor estimate of the logit revenue, but as θ increases the model's optimal value gets closer to the logit optimal value, otherwise always providing an upper bound.

All other experiments presented up to Section 3.7 use $\theta = 0.5$. This yields over 80 per cent of the demand assigned to a shortest path at the optimal solution found by any of the algorithm we have implemented, for any network instance we have tested.

Table 3.1 compares the deterministic heuristic and both a gradient-based ascent method initiated at the origin and a random-start gradient-based method involving 100 replications, with tolls uniformly distributed between 0 and 10. None of these resolution schemes took more than a few seconds to complete. Columns are labelled as follows:

origin	gradient ascent algorithm started at the origin;
rand	100 random-start local searches;
#	number of distinct solution values found by the random-start local searches;
mip	model solution value;
logit	first phase solution value (logit revenue);
+local	second phase solution value (logit revenue);
%gain	improvement brought about by the second phase (percentage);
error	absolute error committed by the model;
shortest	demand assigned to shortest paths by the best solution (percentage).

While the local search initiated from the origin seldom identifies the best solution,

no.	origin	rand	(#)	mip	logit	+local	%gain	error	shortest
A-5-1	6675.35	6898.14	(3)	8006.00	2482.91	6680.98	169.10	5523.09	90.89
B-5-1	124.42	210.38	(5)	107.30	–1828.69	137.60	92.47	1935.99	86.17
A-10-1	8113.40	9858.76	(36)	11648.00	7502.06	9803.50	30.68	4145.94	90.00
A-10-2	7732.52	14118.90	(34)	16920.00	9744.35	14194.90	45.67	7175.65	87.10
B-10-1	61.93	168.22	(21)	63.27	–3416.01	27.29	99.20	3479.28	100.00
B-10-2	252.02	285.70	(66)	302.49	–2438.17	88.24	96.38	2740.65	100.00

Table 3.1: Deterministic-based heuristic and random-start searches.

the random-start strategy does so on almost all instances. The deterministic heuristic performs better than the local search from the origin on the A-instances, but not on the B-instances. The approximation error is significantly larger on the B-instances: indeed, the corresponding values in column "logit" are even negative!

Table 3.2 provides detailed results obtained on instances A-5-2. Each row is associated with a toll arc. Columns are labelled as follows:

arc	toll arc number;
$x_{\text{det}}(t_{\text{det}})$	first phase deterministic assignment;
t_{det}	first phase solution tolls;
$t_{\text{det+}}$	second phase solution tolls;
$x_{\text{logit}}(t_{\text{det}})$	first phase logit assignment;
$x_{\text{logit}}(t_{\text{det+}})$	second phase logit assignment.

We observe, and this example is typical, that while there can exist a large discrepancy between the deterministic and true flow assignments, the deterministic model reaches an optimal or near-optimal concavity region, from where a local search yields a near-optimal solution.

arc	t_{det}	$t_{\text{det+}}$	$x_{\text{det}}(t_{\text{det}})$	$x_{\text{logit}}(t_{\text{det}})$	$x_{\text{logit}}(t_{\text{det+}})$
1	138	133.12	23	11.50	22.83
2	–22	–1.45	0	5.33	0.00
3	128	123.19	16	10.67	15.87
4	40	36.56	58	19.33	49.69
5	38	35.56	0	19.33	6.72

Table 3.2: First phase and second phase solutions on Instance A-5-2

3.5 A bilevel trust-region approach

The deterministic heuristic introduced above relies on a very coarse "all or nothing" approximation of the logit probabilities. In this section, we propose to refine this approximation and cast it within the framework of a trust-region globalization strategy, thus providing an improved descent algorithm. To this end, we replace the nonlinear terms involved in the expression of the logit probabilities by either a discontinuous (stepwise) or continuous (affine) linear approximation. This approach is akin to the bilevel trust-region framework introduced by Marcotte et al. in [57] for general bilevel programs, and implemented and tested by Colson et al. [26] in a variety of situations, including toll setting.

More specifically, let us consider the mathematical program obtained after performing a second-order local approximation of the entropic term in the exact logit bilevel formulation (Program 3.5). This yields a mixed integer program reminiscent of the deterministic formulation (Program 3.1), but whose lower level optimality conditions involve a *linear* approximation of the logarithmic function. This combinatorial optimization model contrasts with the traditional quadratic trust-region models and is better suited to the structure of the pricing problem. We conjecture that, more than improving the ability of the algorithm to converge to a stationary point, the bilevel trust-region framework increases the ability of the ascent algorithm to uncover a good optimum in a global sense.

The trust-region strategy requires repeatedly evaluating the quality of the model. This can be expensive in a bilevel context but, in our case, the logit closed form expression (the lower level solution) makes this evaluation straightforward. Under most toll policies and for most network instances, a significant proportion of the paths carry very small flows. Since the lower level Lagrangian of Program 3.5 is unbounded at the origin, deriving reliable local information close to the origin is numerically challenging.

For $k > 1$ let (t^k, x^k) be the k-th iterate. Assume that $t^k \in T^k$, the associated trust-region defined as an hypercube containing the previous upper level iterate:

$$t^k \in T^k = T \cap \left[t^{k-1} - \frac{\delta_{k-1}}{2} 1_{\mathscr{R}}, t^{k-1} + \frac{\delta_{k-1}}{2} 1_{\mathscr{R}} \right], \tag{3.17}$$

where δ_{k-1} denotes the trust-region radius, which is updated according to the "fitness" of the model. The bilevel model, which provides a local approximation of the logit revenue function, is defined, at iteration k, on the basis of a second order local approximation of g, the lower level objective of Program 3.5, around a point $\hat{x}^k \in X$ close to the exact lower level response to the previous iteration optimal tolls $\hat{x}^k \sim \text{logit}(t^{k-1})$, but should not have components "too close" to zero.

Let us denote $m^{\log x}(\cdot|\hat{x}) : \mathbb{R}^R \to \mathbb{R}^R$ the vector function such that each component $m_r^{\log x}(\cdot|\hat{x})$ is a first-order local approximation of the function $\log x$ around $\hat{x}_r$:

$$m_r^{\log x}(x|\hat{x}) = (\hat{x}_r)^{-1} x_r + \log \hat{x}_r - 1. \tag{3.18}$$

An improved second-order approximation of the lower level objective of Program 3.5 around $\hat{x}^k$ takes the form

$$\begin{aligned} g(t,x) &\approx (t,x) \cdot \nabla_{xx} g(t^k, \hat{x}^k)(t,x) + \nabla_x g(t^k, \hat{x}^k) \cdot (t,x) + g(t^k, \hat{x}^k) \\ &= (c+t) \cdot x + [(m_r^{\log x}(x|\hat{x}^k) + \log \hat{x}^k) \cdot x]/2\theta + \text{cte}, \end{aligned}$$

where constant terms (with respect to x) have been ignored. Since this quadratic form is negative definite, the corresponding quadratic approximation of the lower level of Program 3.5 can be replaced by its first-order optimality conditions to yield the linear complementarity system

$$0 \le c + t + \theta^{-1} m^{\log x}(x|\hat{x}^k) - 1_{\mathscr{Q}}^{\mathscr{R}} \pi \perp x \in X, \tag{3.19}$$

where we have used the fact that logit probabilities are invariant under disutility translation. Next, we linearize the bilinear objective f. For that purpose, two schemes are considered. Colson et al. use in [26] the first-order local approximation:

$$f_m^1(t,x) \approx \sum_{r\in\mathscr{R}} d_{q(r)}\,(\hat{x}_r^k t_r + x_r t_r^k). \tag{3.20}$$

Alternatively, an *exact* concave quadratic expression can also be derived. Indeed, since the lower level variable x satisfies the complementarity system (3.19), we have:

$$f_m^2(t,x) \quad \approx \quad -\sum_{r\in\mathscr{R}} d_{q(r)}[c_r + m_r^{\log x}(x|\hat{x}^k)]\,x_r + \sum_{q\in\mathscr{Q}} d_q \pi_q \tag{3.21}$$

The corresponding models are expressed as:

Program 3.8.

$$\begin{aligned} \max_{t,x,\pi} \quad & f_m^i(t,x) \\ \text{s.t. } & t \in T^k \text{ and } (3.19), \end{aligned}$$

for $i = 1,2$.

The models are solved exactly through the use of an equivalent mixed integer representation. The latter relies, as did the mixed integer program associated with the deterministic toll setting problem, on a "big M" scheme.

Program 3.9.

$$\begin{aligned} \max_{\pi,t,x,z} \quad & f_m^i(t^k,x^k) \\ \text{s.t.} \quad & 0 \le c + t + \theta^{-1} m^{\log x}(x|\hat{x}^k) - 1_{\mathscr{Q}}^{\mathscr{R}}\pi \le M z && (3.22) \\ & x \le 1_{\mathscr{R}} - z && (3.23) \\ & t \in T^k, x \in X, z \in \{0,1\}^R. \end{aligned}$$

Technically, the models are solved until a tolerance level ("gap") is achieved. The computational effort required to solve this program, routinely more than once at each iteration, largely depends on the quality of the associated linear relaxation, and in turn, on the magnitude of the constant M. For notational simplicity, we will assume the presence of a single origin-destination couple. For linear constraints (3.22)–(3.23) to implement complementarity constraint (3.19), the value of M must exceed the largest possible value of the expression in the middle of (3.22), for any feasible x, x^k, t and π. For each $q \in \mathscr{Q}$, the variable π_q is feasible if it is set to the smallest value of $c_r + t_r + \theta^{-1} m_r^{\log x}(x|\hat{x}^k)$ as r runs over $\mathscr{R}^q$. Thus

$$\begin{aligned} M &\geq |c_r + t_r + \theta^{-1}(x_r/\hat{x}_r^k + \log \hat{x}_r^k) - c_\ell + t_\ell + (x_\ell/\hat{x}_\ell^k + \log \hat{x}_\ell^k)/\theta| \\ &= c_r - c_\ell + t_r - t_\ell + (x_r/\hat{x}_r^k - x_\ell/\hat{x}_\ell^k)/\theta + \log(\hat{x}_r^k/\hat{x}_\ell^k) \end{aligned}$$

is a theoretically suitable value, and we can thus set $M = \operatorname{diam} c + \delta_k + 1/\delta_{\min} + \log \delta_{\min}$, where $\operatorname{diam} c$ is defined in (3.8), $\delta_{\min}$ is defined in (3.9) and δ_k is the trust-region radius at iteration k defined in (3.17). A smaller value for M can be derived, but in any case we must ensure that $M > (\hat{x}_r^k)^{-1}, \forall r \in \mathscr{R}$.

We now provide an overview of the sequential optimization process embedded into trust-region strategies, where the ascent phase can be started at any feasible point. Let

- $f_m(t^k, x^k)$: value of the model under toll policy t^k given the lower level approximation is taken around $\hat{x}^{k-1}$;
- $f^{\text{logit}}(t^k)$: value of the exact logit revenue function under toll policy t^k.

An optimal solution of the model is either accepted or rejected, depending on the following goodness of fit measure:

$$\frac{f_m^i(t^{k+1}, x^{k+1}) - f_m^i(t^k, x^k)}{f^{\text{logit}}(t^{k+1}) - f^{\text{logit}}(t^k)}. \tag{3.24}$$

In the above, the numerator measures the improvement brought about by the k-th iterate to the model's optimal value, and the denominator, the corresponding improvement in the logit revenue function. This ratio is larger than one (respectively smaller than one) if an improvement in the model's objective value leads to an improvement of greater magnitude (respectively smaller magnitude) in the exact logit revenue. Depending on this ratio, an iterate is either accepted or rejected. In the latter case, and if the trust radius δ_k is larger than some predetermined threshold, the trust radius δ_k is reduced by some predetermined factor. The rate at which the trust-region radius is either increased of decreased, and the relative tolerance to poor performance before accepting an iterate, are standard parameters in trust-region algorithms. Good results are obtained using standard values. We refer to a specialized work for further details [29].

Throughout our numerical experiments, the vector $\hat{x}^k = \text{logit}(t^{k-1})$ involved components very close to $\delta_{\min}$, resulting in very large values of the parameter M. In this situation Program 3.9 becomes computationally "stiff", and proper measures must be taken so as to insure that no iterate gets too close to zero. This can be achieved in a number of ways, but may have the undesirable effect of halting the algorithm before it reaches a stationary point. However, the situation is less serious than it appears, as the aim of the method is to reach a promising region of concavity, and that "fine tuning" will be left to a subsequent local search.

To bound away from zero the iterates, we considered several options. Setting lower bounds on the flow variables at the upper level may lead to infeasible problems, while setting them at the lower level may yield a solution that differs sharply from a logit assignment, and to premature termination before reaching a good concavity region. While one can define updating rules of the trust-region parameters that succeed in keeping the next lower level solution away from the origin, based on the values of the previous iterates, simpler methods perform as well, and we introduce two such schemes. The first scheme consists in setting the lower level local approximation not too close to the origin, that is

$$\hat{x}^k = \max\{x^{k-1}, \eta 1_{\mathscr{R}}\}, \tag{3.25}$$

where η is set to any appropriate value. This scheme is referred to as the A.L. or A.Q. variant of the bilevel trust-region algorithm, whether it is implemented with a linear upper level objective (3.20), or a quadratic upper level objective (3.21), respectively. The second scheme consists in adaptively dropping paths from $\mathscr{R}$ associated with small choice probabilities. This involves replacing the lower level feasible set X, at iteration k, by X^k such that

$$X^k = \{x| \sum_{r \in \mathscr{R}^k \cap \mathscr{R}^q} x_r = 1,\, q \in \mathscr{Q}\} \tag{3.26}$$

where $\mathscr{R}^k = \{r \in \mathscr{R} | x_r^k \geq \eta\}$ and η is a small integer. This scheme is referred to as the D.L. or D.Q. variant, whether it is implemented, again, with a linear or quadratic upper level objective, respectively. To summarize, we have the four algorithmic variants:

variant A.L.	lower level approximation around $\hat{x}^k$ with a first-order approximation of the objective;
variant A.Q.	lower level approximation around $\hat{x}^k$ with a second-order approximation of the objective;
variant D.L.	lower level feasible set X^k with a first-order approximation of the objective;
variant D.Q.	lower level feasible set X^k with a second-order approximation of the objective.

Next, a trade-off must be achieved between a small value of η (which results in large running times) and a large value of η that may yield a bad approximation of the logit flow. In the first case, the implementation of the lower level complementarity constraints will involve a large big M. In the second case, the estimation of small logit probabilities, which are likely to arise, will not be acurate. We used:

$$\eta = (10R)^{-\gamma}, \tag{3.27}$$

where γ is a small integer value. If the parameter γ is small, the iterates stay away from zero and the resulting sub-problems are quickly solved. However, in this case, the algorithm usually halts before reaching the optimal concavity region, especially for instances where θ is large, most paths then being assigned very small choice probabilities. Using a large value of γ let the iterates get closer to zero, at the cost of stiffer sub-problems.

The four variants of the trust-region algorithm are compared in Table 3.3, where the Phase II solution values are tabulated. The tolerance factor (gap) is set to 10 per cent. Other parameter settings are $\theta = 1$, $\gamma = 2$ and $\delta_0 = 5$. Using different values of δ_0 did not alter significantly the performance. Differences in the solutions can mostly be observed on the B-instances. While variant A.Q. performs well, none of the variant clearly dominates.

A detailed account of the results obtained on instance B-5-2 is given in Table 3.4, where naming conventions for columns are the same as in previous tables, with the exception of column *time* and *iter*, giving respectively the cpu time and the iteration number taken by the trust-region phase. In this case, variants A.L. and A.Q. outperform the other two. All variants are associated with very large approximation errors, as occur on most B-instances. While variant D.Q. achieves the highest revenue (68.92), we note that it does *not* achieve the smallest approximation error or gain, which is achieved by A.Q., in spite of a lower predicted objective equal to 50.29.

We now turn our attention to the A.Q. variant, with Table 3.5 providing a detailed account of its performance. The error committed by the model is always significant, but even so, the gain obtained from the second phase is less than one percent on the A-instances. This statistic is larger on the B-instances, for which the model's final solution is never even positive! Nonetheless, the corresponding logit revenue (column *logit*) is always positive, which was not the case in Table 3.1 for the deterministic heuristic.

no.	D.L.	D.Q.	A.L	A.Q.
A-5-1	**5803.71**	**5803.71**	**5803.71**	**5803.71**
B-5-1	210.32	**210.38**	157.44	210.32
A-10-1	**9858.76**	**9858.76**	**9858.76**	**9858.76**
A-10-2	**12245.50**	12195.80	12207.80	12207.80
B-10-1	142.34	**168.22**	**168.22**	142.34
B-10-2	285.45	**292.02**	274.90	**292.02**

Table 3.3: Bilevel trust region variants: second phase solution values.

algo	time	iter	mip	logit	+local	error	%gain
D.L.	0.42	21	–600.45	17.39	50.29	617.84	189.10
D.Q.	0.39	14	–597.13	18.45	**68.92**	615.58	273.42
A.L.	219.35	500	–560.51	39.94	50.29	600.44	25.93
A.Q.	172.44	443	–550.65	**50.15**	50.29	600.80	0.28

Table 3.4: Bilevel trust region variants: detailed results on Instance B-5-2.

Finally, Table 3.6 contrasts A.Q. against local searches. Columns are organized as follows:

origin	gradient-based search from the origin;
det+local	deterministic heurisitic second phase solution value;
tr+local	trust-region second phase solution value;
det+tr+loc	deterministic warm-started trust-region second phase solution value.

The initial trust-region radius has been set to $\delta_0 = 1$ for the *det+tr+loc* scheme. This value, which is smaller than in the other experiments, aims to help the algorithm better exploit the initial iterate. Comparing columns *origin* and *tr+local*, we observe that the trust-region strategy performs significantly better than a gradient ascent. Compared to the deterministic heuristic, the trust region algorithm finds better solution on the B-instances, with reverse conclusions on the A-instances. However, warm-starting the trust-region algorithm with the deterministic optimal tolls is disappointing, this strategy being outperformed by the deterministic heuristic. The trust-region algorithm actually misses the very concavity region from which it is initiated.

3.6 A piecewise constant approximation scheme

In this section, we consider improved approximations to the original model that will (hopefully) yield a better starting point from which initiating the local ascent phase. These approximations involve the linearization of the lower level entropic term associ-

no.	time	iter	mip	logit	+local	error	%gain
A-5-1	0.64	13	3662.14	5801.67	5803.71	2139.53	0.03
B-5-1	2.19	51	−1534.02	111.76	210.31	1645.78	88.19
A-10-1	3.51	24	7367.81	9858.18	9858.76	2490.37	0
A-10-2	4.10	28	9261.93	12189.50	12207.80	2927.57	0.15
B-10-1	5.28	38	−4482.47	138.97	142.34	4621.44	2.42
B-10-2	13.40	83	−2135.13	233.47	292.02	2368.60	25.08

Table 3.5: A.Q. bilevel trust region: detailed results.

no.	origin	det+local	det+tr+local	tr+local
A-5-1	6675.35	6680.98	**6816.18**	5803.71
B-5-1	124.42	137.60	104.60	**210.32**
A-10-1	8113.40	**10930.30**	5969.97	9858.76
A-10-2	7732.52	**15233.10**	10718.70	12207.80
B-10-1	61.93	27.88	62.78	**142.34**
B-10-2	252.02	89.13	243.19	**292.02**

Table 3.6: Warm-started A.Q. bilevel trust region: detailed results.

ated with the logit bilevel formulation (Program 3.5). This approximation can be made arbitrarily accurate, provided that sufficient computational power is available. The question of interest is how further refinements of the combinatorial model, and thereby of the underlying logit flows, impact the final solution obtained at the end of the ascent phase.

Let us consider piecewise linear approximations of the entropic term in Program 3.5 or, equivalently, a step function approximation of the sole nonlinear terms of Program 3.5's lower level system, with the exception of the logical complementarity constraints, which can be linearized through the introduction of binary variables.

On the theoretical side, we provide asymptotic bounds on the error, as the number of steps grows. On the computational side, the current scheme involves one extra complementarity constraint per path per constant piece. Increasing the number of constant pieces quickly gets expensive, even on small instances. Whenever the number of constant segments gets large, a heuristic procedure for addressing the model's resolution is required, and such procedure will be proposed and analyzed.

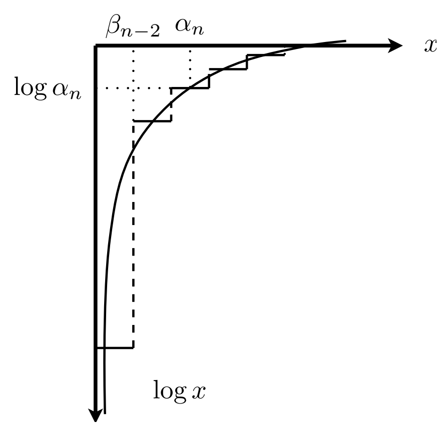


Figure 3.6: Logarithm step function approximation.

The combinatorial approximation is derived as follows. Let $(\alpha_n)_{n\in\mathscr{N}}$ be such that $0 < \alpha_n < \alpha_m \leq 1, \forall n,m \in \mathscr{N}$ such that $n < m$. A tangent to $x\log x$ is defined at each point of this sequence, and we let $m^{x\log x} : \mathbb{R}^R \to \mathbb{R}^R$ such that

$$\begin{aligned} m_r^{x\log x}(x \mid \alpha_n) &= (\log\alpha_n + 1)(x_r - \alpha_n) + \alpha_n \log\alpha_n \\ &= (\log\alpha_n + 1)x_r - \alpha_n\,, \end{aligned} \tag{3.28}$$

for all $r \in \mathscr{R}$. This yields the mathematical program:

Program 3.10.

$$\max_{t,x,w} \quad \sum_{r\in\mathscr{R}} \quad f(t,x) = \sum_{q\in\mathscr{Q}} d_q t^q \cdot x^q \tag{3.29}$$

$$\begin{aligned} \text{s.t.} \quad & t \in T \\ & \min_{w,x} \quad (c+t)\cdot x + \theta^{-1} 1_{\mathscr{R}} \cdot w \\ & \quad s.t. \quad w \geq m^{x\cdot\log x}(x \mid \alpha_n) \qquad n \in \mathscr{N} \qquad (\varphi) \qquad (3.30) \\ & \qquad\qquad x \in X \qquad\qquad\qquad\qquad\qquad (\pi), \end{aligned}$$

where φ is the vector of multipliers associated with the linear constraints (3.30) and π is the vector of multipliers associated with the simplex constraints in the definition of the lower level feasible set X. Constraint (3.30) restricts each component of w to an intersection of N half-spaces above the tangents to the convex function $x\log x$. Minimization at the lower level insures that $1_{\mathscr{R}} \cdot w$ behaves like a piecewise linear approximation of $x\log x$. The approximation can be made arbitrarily precise over $X_0 = \{x | x \geq \delta_{\min} 1_{\mathscr{R}}\}$, where $\delta_{\min}$ is defined in (3.9), given that the sequence $\alpha^{x\log x}$ meets appropriate requirements. Let (t,x,w,π,φ) be a KKT point and assume that the points in $\alpha^{x\log x}$ are equidistant. With reference to the exact bilevel formulation lower level objective g (Program 3.5) we immediately have

$$(c+t)\cdot x + \theta^{-1} 1_{\mathscr{R}} \cdot w = g(t,x) + O(N^{-2}).$$

Also, on account of the nature of Program 3.10's lower level, the upper level bilinear expression f is equivalently expressed as a concave quadratic function. Indeed strong duality at the lower level states that

$$(c^q + t^q) \cdot x^q + \frac{1}{\theta} 1_{\mathscr{R}^q} \cdot w^q = -\frac{1}{\theta} \sum_{n \in \mathscr{N}} \sum_{r \in \mathscr{R}^q} \varphi_{rn} \, \alpha_n + \pi_q,$$

for each $q \in \mathscr{Q}$, where φ_{rn} is the Lagrange multiplier of constraint (3.30) associated with path r and the n-th tangent $m_r^{x\log x}(\cdot|\alpha_n)$. Isolating $t^q \cdot x^q$ for each $q \in \mathscr{Q}$ yields

$$f(t,x) = \sum_{r \in \mathscr{R}} d_{q(r)} [-c_r x_r - \frac{1}{\theta} w_r - \frac{1}{\theta} \sum_{n \in \mathscr{N}} \alpha_n \varphi_{rn} + \pi_{q(r)}]. \tag{3.31}$$

In practice, we implemented the combinatorial model using a slightly different objective, the rationale for this decision being as follows. Positive entries of the multiplier vector $\varphi_{r\cdot}$ are associated with linear segments $m_r^{x\log x}(x|\cdot)$ active in the piecewise linear approximation of $x_r \log x_r$, i.e., that correspond to active constraints in (3.30). At most two such entries are positive, they have consecutive indices and sum to 1. The positivity of $\varphi_{rn} > 0$ implies that $x_r \log x_r$ is approximated by a tangent taken at $\alpha_n^{x\log x}$, which in turn implies that x_r lies close to $\alpha_n^{x\log x}$. Thus for any given path $r \in \mathscr{R}^q$ and given that sequence $\alpha^{x\log x}$ satisfies minimal requirements (equidistance, for instance), we have that $w_r \to x_r \log x_r$ as $N \to \infty$, which implies $\sum_{n=1}^{N} \varphi_{rn} \alpha_n^{x\log x} \to x_r$ and thus $\sum_{r \in \mathscr{R}^q} \sum_{n=1}^{N} \varphi_{rn} \alpha_n^{x\log x} \to 1$. This motivates the use of the simpler objective

$$f(t,x) \sim \sum_{r \in \mathscr{R}} d_{q(r)} [-c_r x_r - \frac{1}{\theta} w_r + \pi_{q(r)}] - \sum_{q \in \mathscr{Q}} d_q / \theta. \tag{3.32}$$

Dropping the constant term yields the single level formulation

Program 3.11.

$$\max_{x,w,t} \quad \sum_{r\in\mathscr{R}} d_{q(r)}[-c_r x_r - \frac{1}{\theta} w_r + \pi_{q(r)}] \tag{3.33}$$

$$\text{s.t.} \quad t \in T, x \in X, \tag{3.34}$$

$$\varphi_{\cdot n} \in X, \qquad n \in \mathscr{N},$$

$$0 \leq w - m^{x\log x}(x|\alpha_n^{x\log x}) \quad \perp \quad \varphi_{\cdot n} \geq 0, \qquad n \in \mathscr{N}, \tag{3.35}$$

$$0 \leq c + t + \frac{1}{\theta}\sum_n (\log \alpha_n^{x\log x} + 1_{\mathscr{R}})\, \varphi_{\cdot n} - 1_{\mathscr{Q}}^{\mathscr{R}} \pi \quad \perp \quad x \geq 0, \tag{3.36}$$

where $\log \alpha_n^{x\log x} + 1_{\mathscr{R}} = \nabla_x m^{x\log x}(x|\alpha_n^{x\log x})$. In the above, constraint (3.35) ensures that w is a piecewise linear approximation of the function $x\log x$. The associated multiplier φ allows to define the step function approximation of the logarithm in constraint (3.36).

A variational formulation is helpful in characterizing the approximation error. To this aim, we introduce the sequence of points $(\beta_n^{x\log x})_{n=1}^{N-1}$ at which each tangent $m_r^{x\log x}(x|\alpha_n)$ and $m_r^{x\log x}(x|\alpha_{n+1})$ meet

$$m^{x\log x}(\beta_n^{x\log x}|\alpha_n) = m^{x\log x}(\beta_n^{x\log x}|\alpha_{n+1}), \quad 1 \leq n \leq N-1 \tag{3.37}$$

and consider the program

Program 3.12.

$$\max_{x,t} \quad f(x,t)$$

$$\text{s.t.} \quad t \in T$$

$$x \in \text{SOL}(F^0(t,\cdot|\alpha), X), \tag{3.38}$$

where

$$F_r^0(t,x|\alpha) = \begin{cases} F_r(t,\alpha_n 1_{\mathscr{R}}) & \text{if } x_r \in (\beta_{n-1}^{x\log x}, \beta_n^{x\log x}) \\ k \in [F_r(t,\alpha_n 1_{\mathscr{R}}), F_r(t,\alpha_{n+1} 1_{\mathscr{R}})] & \text{if } x_r = \beta_n^{x\log x}. \end{cases}$$

The following lemma characterizes the piecewise linear approximation used in Program 3.10 and allows to show the equivalence between Program 3.10 and Program 3.12.

Lemma 3.2. *Let w and x be feasible for Program 3.10. Then w is a piecewise linear function of x such that*

$$w = \sum_{n=1}^{N} m^{x\log x}(x|\alpha_n)\mathbf{1}\{x \in (\beta_{n-1}^{x\log x}, \beta_n^{x\log x}]\},$$

where $\mathbf{1}\{\cdot\}$ is the indicator function.

Proof. The convexity of $x\log x$ ensures that $\beta^{x\log x}$ is an increasing sequence such that, for all r, there exists n such that

$$x_r \in (\beta_{n-1}, \beta_n] \Rightarrow m_r^{x\log x}(x \mid \alpha_n) \geq m_r^{x\log x}(x \mid \alpha_\ell)$$

for all $\ell = 1, \ldots, N$. The lower level of Program 3.10 is bounded, and at least one of the inequality constraints (3.30) associated with one of the linear segments $m^{x\log x}(x|\cdot)$ and index r is tight. More precisely $w_r = m_r^{x\log x}(x \mid \alpha_n)$, and the result follows. □

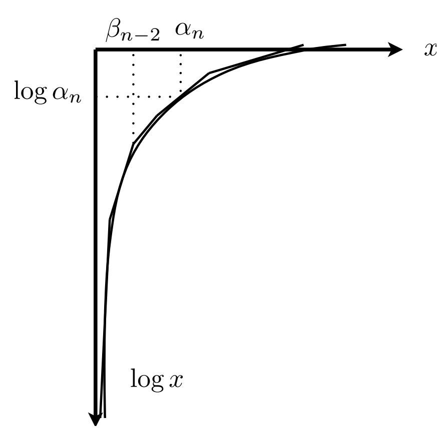


Figure 3.7: Piecewise linear approximation of the logarithmic function.

Theorem 3.4. *Programs 3.10 and Program 3.12 are equivalent.*

Proof. Let $\tilde{g}$ denote the lower level objective of Program 3.10. We show that F^0 is a subgradient of $\tilde{g}$ with respect to t. Let (x,w,t) be a feasible solution. It is a consequence of Lemma 3.2 that $\tilde{g}(x,t,w)$, as a function of x, is composed of the linear segments $\{(c+t)\cdot x+\theta^{-1}m^{x\log x}(x|\alpha_n)\}_{n\in\mathscr{N}}$. Then, for all $x\in X$, there exists $n\in\mathscr{N}$ such that

$$x\in(\beta_{n-1},\beta_n]\Rightarrow\nabla_x\tilde{g}(t,x,w)=c+t+\theta^{-1}\nabla_x m^{x\log x}(x|\alpha_n)=F(t,\alpha_n 1_{\mathscr{R}}).$$

The conclusion follows. □

Theorem 3.5. *Let (t,x) be feasible for the exact formulation of Program 3.6, and (t,x^0) be feasible for the step function formulation (Program 3.12). Assume that the sequence $\alpha^{x\log x}$ is such that $\dim\alpha^{x\log x}=N$ and that the sequence $\beta^{x\log x}$, defined in (3.37), is made of equidistant points, that is, the step function approximation of the logarithm is made of $N-1$ steps of equal width. Then $f(t,x^0)=f(t,x)+O(1/N)$.*

Proof. The proof is provided for a single OD pair, the generalization to multiple OD pairs being straightforward. Let $n(r)$ such that $x_r^0\in(\beta_{n(r)-1}^{x\log x},\beta_{n(r)}^{x\log x}]$. From Lemma 3.2 we have that the vector function $F^0(\cdot,\cdot|\alpha)$ is constant over $(\beta_{n-1}^{x\log x},\beta_n^{x\log x})$, and so

$$|F_r(t,x^0)-F_r^0(t,x^0|\alpha)|\leq(\theta\delta_{\min})^{-1}|x_r-\alpha_n|\leq(\theta\delta_{\min})^{-1}(\beta_{n(r)}^{x\log x}-\beta_{n(r)-1}^{x\log x})\leq(N\theta\delta_{\min})^{-1},$$

where $(\theta\delta_{\min})^{-1}$ is the largest eigenvalue value of F's Jacobian. Thus

$$|F(t,x^0)-F^0(t,x^0|\alpha)|=\sqrt{\sum_r(F_r(t,x^0)-F_r^0(t,x^0|\alpha))^2}\leq\sqrt{R}/(N\theta\delta_{\min})$$

and the conclusion follows from Lemma 3.1. □

Using a "big M" scheme to implement complementarity constraints (3.35) and (3.36) yields the mixed integer program:

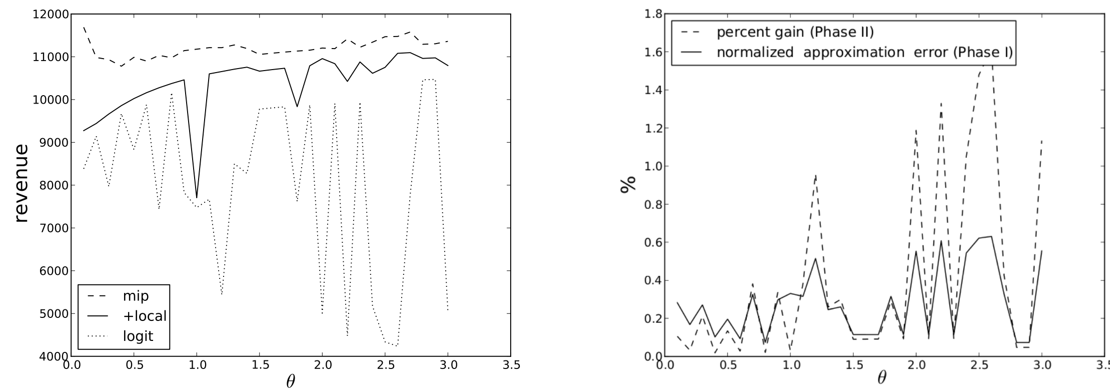


Figure 3.8: Impact of θ on the piecewise constant scheme (instance A-10-1): Phase I and Phase II solution values (left); normalized approximation error and percentage gained from Phase II (right).

Program 3.13.

$$\max_{x,t,\pi\varphi,z^1,z^2} \quad \sum_{q\in\mathscr{Q}} d_q \left(-c^q \cdot x^q - \frac{1}{\theta} 1_{\mathscr{R}^q} \cdot w^q + \pi_q \right)$$

$$\text{s.t.} \quad 0 \leq w - \log \alpha_n^{x\log x} x + 1_{\mathscr{R}}\, \alpha_n \leq M^1 z^1 \tag{3.39}$$

$$0 \leq -\frac{1}{\theta} \sum_n (\log \alpha_n^{x\log x} + 1_{\mathscr{R}})\, \varphi_{\cdot n} + 1_{\mathscr{Q}}^{\mathscr{R}} \pi - c - t \leq M^2 z^2 \tag{3.40}$$

$$\varphi_{\cdot n} \leq 1_{\mathscr{R}} - z^1 \tag{3.41}$$

$$x \leq 1_{\mathscr{R}} - z^2 \tag{3.42}$$

$$x, \varphi_{\cdot n} \in X, t \in T, z^1 \in \{0,1\}^{R\times N}, z^2 \in \{0,1\}^R.$$

In our numerical experiments, good results were obtained by setting

$$\alpha_n^{x\log x} = (10R)^{-\frac{\gamma(N-n-1)}{N-1}},\ 1 < n < N, \tag{3.43}$$

where constraints (3.39) and (3.41) implement the complementarity constraint (3.35), while constraints (3.40) and (3.42) implement the complementarity constraint (3.36). The constant M^1 must be set to a value at least equal to the largest possible value of the expression in the middle of (3.39), for every point in $\alpha^{x\log x}$, as x runs over X and w is feasible. For each $r \in \mathscr{R}$, variable w_r is present in all components of constraint (3.39) with indices (r,n) for $n \in \mathscr{N}$, and so it is equal to the smallest value of $\log \alpha_n^{x_r \log x_r} x_r - \alpha_n$ as n runs over $\mathscr{N}$. Thus

$$M^1 \quad \geq \quad \max\{[\log \alpha_n^{x_r \log x} - \log \alpha_m^{x_r \log x}]x + \alpha_n^{x_r \log x} - \alpha_m^{x_r \log x} \,|\, 0 \leq x \leq 1, 1 \leq n,m \leq N\}.$$

We can set $M^1 = -\log \alpha_1^{x\log x} + 1$. For each $q \in \mathscr{Q}$, the variable π_q occurs in all components of constraint (3.40) with indices (r,n) for $r \in \mathscr{R}^q$ and $n \in \mathscr{N}$ and so it is equal to the smallest value of $\frac{1}{\theta}\sum_n(\log \alpha_n^{x\log x} + 1)\varphi_{rn} + -c_r - t_r$ as r runs over $\mathscr{R}^q$. Thus

$$M^2 \quad \geq \quad \max\{-\frac{1}{\theta}\sum_{n\in\mathscr{N}} \log \alpha_n\,(\varphi_{rn} - \varphi_{\ell n}) - c_r - t_r + c_\ell + t_\ell | r,\ell \in \mathscr{R}, \varphi_{\cdot n} \in X, \forall n, t \in T\}.$$

We can then set M^2 to $-\frac{1}{\theta}\log \alpha_1^{x\log x} + \operatorname{diam} c + \operatorname{diam} T$.

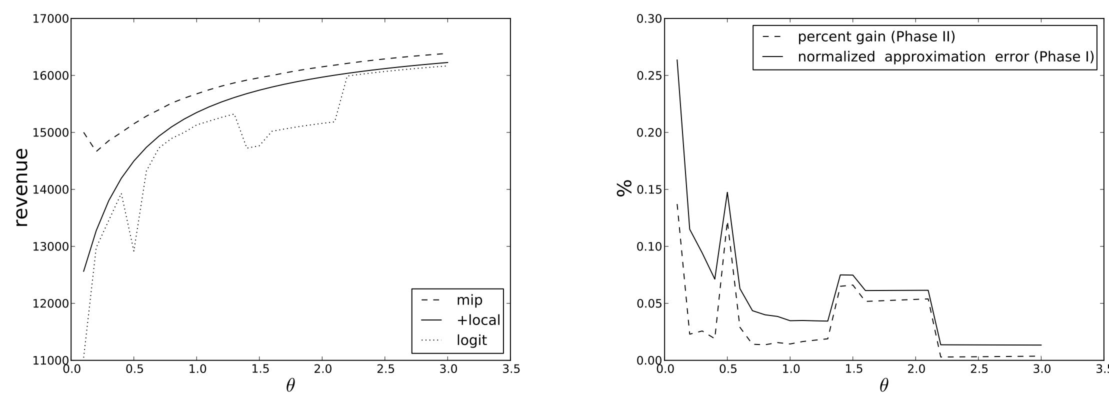


Figure 3.9: Impact of θ on the piecewise constant scheme (instance A-10-2): Phase I and Phase II solution values (left); normalized approximation error and percentage gained from Phase II (right).

The sequence is such that $\alpha_1^{x\log x} = (10R)^{-\gamma}$, $\alpha_N^{x\log x} = 1$, and the logarithm function is approximated by a step function whose steps have equal height. A small value of γ ensures that no points are chosen close to zero, thus avoiding numerical difficulties that mirror those mentioned in the previous section. Also, γ should not be too large, making the mixed integer program difficult to solve, nor too low, which results in a bad approximation of the logit flows, especially for large values of the parameter θ, when logit probabilities are concentrated on shortest paths.

In Figures 3.8 and 3.9 we illustrate the impact of the parameter θ on instances A-10-1 and A-10-2, respectively, in the case of a three-step function. The left-hand side shows the Phase I, Phase II, and the model solution values, respectively. On the right-hand side the corresponding approximation errors and the percentage of gain obtained in Phase II are displayed. We observe that, as θ increases, the three values are close. This contrast with the results obtained with the deterministic heuristic on instance A-10-1 (Figure 3.5), where the deterministic model value did not get any closer to solutions found in Phase I and II using larger values of θ. Still the values achieved in Phase II by the deterministic heuristic and the three-step function approximation are identical. This behavior is common in our numerical experiments, where an improvement in the logit flows approximation does not necessarily yield improved Phase II solutions.

Tables 3.7 and 3.8 provide the Phase II solution values and the corresponding gain increase over Phase I using between 1 and 5 constant segments. CPU times did not exceed 40 seconds for any of these experiments. Note however that using a larger number of segments quickly gets expensive. For instance, solving instance A-10-1 with 7 linear segments takes more than 10 minutes. A quick look at the first table shows that the so-

no.	1	2	3	4	5
A-5-1	6680.98	6680.98	**6898.14**	**6898.14**	**6898.14**
B-5-1	137.60	124.47	**210.38**	**210.38**	**210.38**
A-10-1	9803.50	**9858.76**	**9858.76**	9734.08	**9858.76**
A-10-2	**14194.90**	14194.70	**14194.90**	14194.70	14194.70
B-10-1	27.29	61.45	**168.21**	142.35	**168.21**
B-10-2	88.24	**285.91**	**285.91**	284.65	**285.91**

Table 3.7: Step function based heuristic (1-5 steps): second phase solution values.

no.	1	2	3	4	5
A-5-1	169.08	12.65	3.34	1.02	0.79
B-5-1	107.52	68.28	29.05	29.53	26.34
A-10-1	30.68	22.65	2.47	4.68	3.20
A-10-2	45.67	23.87	1.90	12.59	4.10
B-10-1	100.80	124.36	239.92	82.55	41.31
B-10-2	103.62	17.53	23.18	20.16	0.87

Table 3.8: Step function based heuristic (1-5 steps): improvement brought about by the second phase (percentage).

lution found in Phase II does not improve monotonically with N. On all instances the best solution is actually found with $N \leq 3$. This situation holds true for larger values of N (we tried up to $N = 20$). While the approximation error and the gain from Phase II tend to get smaller with larger values of N, sufficiently so to argue that a global optimum has likely been reached on most of the A-instances, these quantities remain large on the B-instances, for computationally manageable values of N.

Table 3.9 provides a detailed account of the results corresponding to the choice $N = 3$. We observe that the model's optimal value provides a much better approximation of the logit revenue than the deterministic approximation. It also provides an upper bound on both the associated logit revenue and the Phase II solution. We conjecture that a sufficiently small value of θ yields an arbitrarily large Phase II solution as large as

no.	time	mip	logit	+local	%gain	error
A-5-1	0.20	7211.68	6745.22	6898.14	2.28	466.46
B-5-1	0.04	430.05	154.15	210.38	36.48	275.90
A-10-1	53.15	10610.40	9323.04	9734.08	4.41	1287.36
A-10-2	14.80	14818.70	11395.20	14194.90	24.57	3423.50
B-10-1	0.30	502.01	48.76	168.22	245.00	453.25
B-10-2	0.84	401.95	234.77	285.79	21.73	167.18

Table 3.9: 3-step function result details.

desired, while the value of the deterministic optimum remains unchanged.

Table 3.10 presents results obtained after warm-starting local methods with an optimum of the step function approximation scheme. Columns labels are as follows:

det+local	deterministic heuristic second phase solution value;
tr+local	trust-region second phase solution value;
det+tr+local	deterministic warm-started trust-region second phase solution value;
pwc3+local	3-step function scheme second phase solution value;
pwc3+tr+local	3-step function warm-started trust-region second phase solution value.

We observe that the three-step scheme performs better than both the deterministic heuristic and the bilevel trust-region algorithm on this data set.

3.7 A piecewise linear approximation scheme

In this section, we consider piecewise linear (rather than piecewise constant) approximation schemes for the nonlinear terms in the objective and the constraints of Program 3.7. To this aim, let $(\alpha_n^{x\log x})_{n\in\mathcal{N}_{\mathrm{OBJ}}}$ and $(\alpha_n^{\log x})_{n\in\mathcal{N}_{\mathrm{CON}}}$ be positive increasing sequences. Tangents to $x\log x$ and $\log x$ are defined for each point in $\alpha^{x\log x}$ and $\alpha^{\log x}$, respectively. Consider the auxiliary program

no.	det+local	tr+local	det+tr+local	pwc3+local	pwc3+tr+local
A-5-1	6680.98	5803.71	6680.98	**6898.14**	**6898.14**
B-5-1	137.60	210.32	**210.38**	**210.38**	**210.38**
A-10-1	9803.50	**9858.76**	9803.50	**9858.76**	**9858.76**
A-10-2	14194.90	12207.80	10718.70	14194.90	14194.90
B-10-1	27.29	142.34	61.95	**168.21**	79.05
B-10-2	88.24	292.02	**296.07**	285.91	285.91

Table 3.10: Deterministic and step function based warm start strategies for the A.Q. bilevel algorithm: first and second phase solution values.

Program 3.14.

$$\max_{y} \quad \xi \tag{3.44}$$

$$\text{s.t.} \quad \xi \leq m^{\log x}(x|\alpha_n^{\log x}) \qquad (\varphi^n) \qquad n \in \mathscr{N}_{CON}, \tag{3.45}$$

where φ are the multipliers associated with constraints (3.45), which restrict ξ to the intersection of N_{CON} half-spaces below tangents to the logarithm function. Maximizing over ξ ensures that ξ behaves like a piecewise linear approximation *(over-estimation)* of the concave function $\log x$. The current approximation scheme is obtained after replacing $\log x$ in the constraint of Program 3.7 by a variable v and using a piecewise linear approximation of the entropic term, yielding the program

Program 3.15.

$$\max_{t,x,w,v,\pi,\varphi} \quad \sum_{q\in\mathscr{Q}} d_q \left(-c^q \cdot x^q - \frac{1}{\theta} 1_{\mathscr{R}^q} \cdot w^q + \pi_q\right)$$

$$\text{s.t.} \quad t \in T, x \in X \tag{3.46}$$

$$\varphi_{\cdot n} \in X \qquad n \in \mathscr{N}_{\text{CON}}$$

$$w \geq m^{x\log x}(x|\alpha_n^{x\log x}) \qquad n \in \mathscr{N}_{OBJ} \tag{3.47}$$

$$0 \leq v - m^{\log x}(x|\alpha_n^{\log x}) \quad \perp \quad \varphi_n \geq 0 \qquad n \in \mathscr{N}_{CON} \tag{3.48}$$

$$0 \leq c + t + \frac{1}{\theta} v + 1_{\mathscr{Q}}^{\mathscr{R}} \pi \quad \perp \quad x \geq 0, \tag{3.49}$$

where the vector function $\alpha^{x\log x}$ is defined in (3.43), and $m^{\log x}$ is defined in (3.18). The constraint (3.47) restricts each component of w to an intersection of N_{OBJ} half-spaces located above tangents to the convex function $x \log x$. Maximization of the objective ensures that w_r behaves like a piecewise linear approximation (under-estimation) of $x_r \log x_r$, for each $r \in \mathscr{R}$. Also, for each $r \in \mathscr{R}$, the subset of constraints $0 \leq v_r - m_r^{\log x}(x|\alpha_n^{\log x}) \perp \varphi_{rn}$, $n \in \mathscr{N}_{CON}$ in (3.48) corresponds to the first-order optimality conditions of Program 3.14.

The following result provides a closed form expression for v, which will be useful for error estimation purposes.

Lemma 3.3. *Let* (t,x,w,v,π,φ) *a KKT point of Program 18. Define* $(\beta_n^{\log x})_{n=1}^{N-1}$ *as the sequence of points where the tangent couples* $m^{\log x}(x|\alpha_n^{\log x})$ *and* $m^{\log x}(x|\alpha_{n+1}^{\log x})$ *intersect. Then v is a piecewise linear function of x such that*

$$v = \sum_{n=1}^{M} m^{\log x}(x|\alpha_n)\mathbf{1}\{x \in (\beta_{n-1}^{\log x}, \beta_n^{\log x}]\}.$$

Proof. The concavity of $\log x$ ensures that $\beta^{\log x}$ is an increasing sequence such that , for every $r \in \mathscr{R}$, there exists m such that $x_r \in (\beta_{m-1}^{\log x}, \beta_m^{\log x}] \Rightarrow m_r^{\log x}(x \mid \alpha_m) \geq m_r^{\log x}(x \mid \alpha_\ell), 1 \leq \ell \leq M$. Program 3.14 is bounded and at least one of the linear segments $m_r^{\log x}(x|m')$, for $1 \leq m' \leq M$, must be tight in constraint (3.45). More precisely $v_r = m_r^{\log x}(x|\alpha_m)$. □

Now, consider the following program where the choice probability vector x is expressed as a solution of a variational inequality (first-order optimality conditions):

Program 3.16.

$$\max_{t\in T,(x,\pi)\in X} \quad f^1(t,x) = \sum_{q\in\mathscr{Q}} d_q \left(-c^q \cdot x^q - \frac{1}{\theta} 1_{\mathscr{R}^q} \cdot w^q + \pi_q \right) \tag{3.50}$$

$$\begin{aligned} \text{s.t.} \quad & t \in T \\ & w \geq m^{x\log x}(x|\alpha_n^{x\log x}) \qquad n \in \mathscr{N}_{obj} \\ & x \in \text{SOL}(F^1, X), \end{aligned} \tag{3.51}$$

where

$$F^1(t,x) \quad = \quad \sum_{n\in\mathscr{M}} [\nabla_x F(t,\alpha_n^{\log x} 1_{\mathscr{R}})(x - \alpha_n^{\log x} 1_{\mathscr{R}}) + F(t,\alpha_n^{\log x} 1_{\mathscr{R}})] \mathbf{1}\left\{ x \in (\beta_n^{logx}, \beta_{n+1}^{logx}] \right\},$$

with F defined as in (3.11), the exact lower level objective's gradient with respect to x. We have the following result.

Theorem 3.6. *Programs 3.15 and 3.16 are equivalent.*

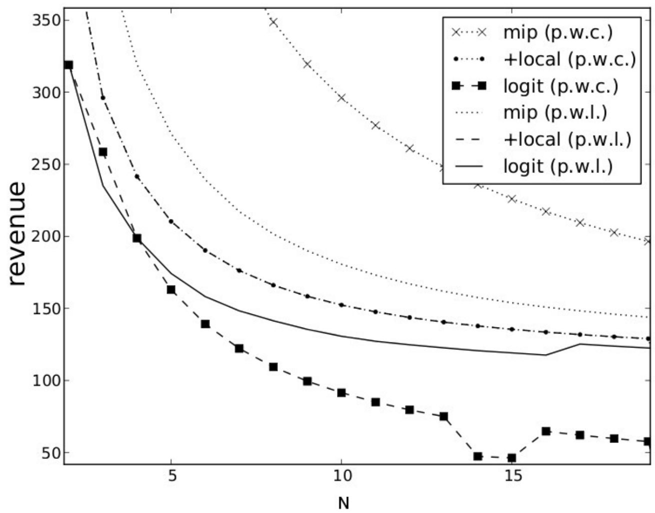


Figure 3.10: Impact of N and N_{CON} (instance B-5-1). Phase I and Phase II for the piecewise constant and piecewise linear schemes.

Proof. The function F^1 is obtained from $F = \nabla_x g$ (lower level objective gradient of the logit bilevel formulation (3.11)) after replacing the logarithm in $F(t,x) = c + t + \theta^{-1}\log(x+1)$ by the piecewise linear approximation:

$$\begin{aligned}
F^1(t,x) &= \sum_{n\in\mathcal{N}} [\nabla_x F(t,\alpha_n 1_{\mathscr{R}})(x-\alpha_n 1_{\mathscr{R}}) + F(t,\alpha_n 1_{\mathscr{R}})] 1\{x \in (\beta_n, \beta_{n+1}]\} \\
&= \sum_{n\in\mathcal{N}} [(\theta\alpha_n)^{-1}(x-\alpha_n 1_{\mathscr{R}}) + c + t + \theta^{-1}(\log\alpha_n 1_{\mathscr{R}} + 1_{\mathscr{R}})]\mathbf{1}\left\{x \in (\beta_n^{\log x}, \beta_{n+1}^{\log x}]\right\} \\
&= \theta^{-1} 1_{\mathscr{R}} + c + t + \theta^{-1} \sum_{n\in\mathcal{N}} m^{\log x}(x|\alpha_n)\mathbf{1}\left\{x \in (\beta_n^{\log x}, \beta_{n+1}^{\log x}]\right\}
\end{aligned}$$

Theorem 3.7. *Let (t,x) be feasible for the exact logit formulation (Program 3.6) and (t,x^1) feasible for the piecewise linear scheme approximate formulation (Program 3.16), given that N_{con} linear segments are used to approximate the logarithm and N_{obj} segments to approximate the entropic term. Assume further that the sequences $\alpha^{\log x}$ and $\alpha^{x\log x}$ are chosen so that the sequences $\beta^{x\log x}$ and $\beta^{\log x}$ are composed of equidistant points. Let $f^1(t,x^1)$ denote the objective of Program 3.16 as a function of the toll and the flow,* *and $f(t,x)$ the objective of Program 3.6. Then*

$$f^1(t,x^1) = f(t,x) + O([\max\{N_{\text{con}}, N_{\text{obj}}\}]^{-2}).$$

Proof. Let F and F^1 be defined as in Program 3.6 and Program 3.16, respectively (lower level gradients). We first show that (i) $F^1(x^1) = F(x^1) + O(1/N_{\text{con}}^2)$. By applying Theorem 3.1 we then get $f(t,x^1) = f(t,x) + O([N_{\text{con}}]^{-2})$. Next we prove that (ii) $f^1(t,x^1) = f(t,x^1) + O([N_{\text{obj}}]^{-2})$, from which the conclusion will follow. (i) For $r \in \mathscr{R}$, let $n(r)$ be such that $x_r^1 \in (\beta_{n(r)-1}^{\log x}, \beta_{n(r)}^{\log x}]$. From Lemma 3.3 it follows that F_r^1 is a first-order approximation of F_r over $(\beta_{n(r)-1}^{\log x}, \beta_{n(r)}^{\log x})$. Thus $|F_r(t,x^1) - F_r^1(t,x^1)| \leq (\theta\delta_{\min})^{-1}(x_r - \alpha_n^{\log x})^2 \leq (\theta\delta_{\min})^{-1}(\beta_{n(r)}^{\log x} - \beta_{n(r)-1}^{\log x})^2 \leq (N^2\theta\delta_{\min})^{-1}$, where $(\theta\delta_{\min})^{-1}$ is the largest eigenvalue value of F's Jacobian, and $\delta_{\min}$ is the smallest logit choice probability (defined as in (3.10). It follows that

$$|F(t,x^1) - F(t,x^1|\alpha^{\log x})| = \sqrt{\sum_r [F_r(t,x^1) - F_r^1(t,x^1)]^2} \leq \sqrt{R}/(N_{\text{con}}^2\theta\delta_{\min})$$

(ii) The function f^1 is derived from a piecewise linear approximation of the entropic term in the exact logit revenue expression of Program 3.7. The result follows from the construction of the sequence $\beta^{x\log x}$, where the width of each step is inversely proportional to N_{con}. □

Since no element of the sequences $\log x$ or $x\log x$ should be close to 0, we based our implementation on (3.43) . Using a "big M" scheme to implement Program 3.15 yields

Program 3.17.

$$
\begin{aligned}
\max \quad & 1_{\mathscr{Q}}^{\mathscr{R}} d \cdot (1_{\mathscr{Q}}^{\mathscr{R}} \pi - c \cdot x - w/\theta) && \\
\text{s.t.} \quad & w - \log \alpha_n^{x \log x} x + \alpha_n^{x \log x} 1_{\mathscr{R}} \leq 0 && n \in \mathscr{N}_{obj} && (3.52)\\
& v - x/\alpha_n^{\log x} - (\log \alpha_n^{\log x} - 1) 1_{\mathscr{R}} \leq M^z z^n && n \in \mathscr{N}_{con} && (3.53)\\
& \varphi_n \leq 1_{\mathscr{R}} - z^n && n \in \mathscr{N}_{con} && (3.54)\\
& \varphi^n \in X && n \in \mathscr{N}_{con} && (3.55)\\
& c + t + \frac{1}{\theta} v - 1_{\mathscr{Q}}^{\mathscr{R}} \pi \leq M^y y && && (3.56)\\
& x \leq 1_{\mathscr{R}} - y && && (3.57)\\
& x \in X && && (3.58)\\
& t \in T && &&\\
& z \in \{0,1\}^{RN}, y \in \{0,1\}^R. && &&
\end{aligned}
$$

In the above program, constraints (3.53)–(3.55) implement the complementarity (3.48), while constraints (3.56)–(3.58) implement the complementarity (3.49). The constant M^z must be set to a value at least equal to the largest possible value of any component r of the expression on the left-hand side of (3.53), for every point in $\alpha^{\log x}$ and $x \in \{0,1\}$. For each $r \in \mathscr{R}$, the variable v_r is present in all components of constraint (3.53) with indices (r,n) for $n \in \mathscr{N}_{con}$:

$$M^z \geq \max\{y/\alpha_n^{\log x} - \log \alpha_n^{\log x} - y/\alpha_m^{\log x} + \log \alpha_m^{\log x} | 0 \leq y \leq 1, 1 \leq n,m \leq N_{\text{con}}\}$$

For instance, one may set $M^z = 2/\alpha_1^{\log x}$. The Constant M^y must be at least equal to the largest value of any component r of the expression on the left-hand side of (3.56), for every point in $\alpha^{\log x}$, x, π and v feasible. Using Lemma 2 this yields

$$
\begin{aligned}
M^y &= \max\{v_r - c_r - t_r + v_\ell + c_\ell + t_\ell \mid r, \ell \in \mathscr{R}\}\\
&= \operatorname{diam} c + \operatorname{diam} T - \log \delta_{\min}.
\end{aligned}
$$

Program 3.15 involves the same number of complementarity constraints as the step function approximation (Program 3.11). However, the resulting mixed integer program is more challenging. For the majority of experiments we set $N_{\text{OBJ}} = 10$. On stiff problems however (ten toll A-instances with $N_{\text{CON}} \geq 3$) we set N_{OBJ} to zero. This has an impact on the solution, but is consistent with the logit flow approximation.

Tables 3.11 and 3.12 show the Phase II objective values and the associated gain, using one to five linear segments. CPU times, which are on some instances significantly larger than those of Section 3.6 , are given in Table 3.13. Comparing these results with those of Section 3.6 (Tables 3.7 and 3.8) we observe that, given an equal number of linear segments, the piecewise linear approximation yields a much better approximation of the logit flow. In this regard the numerical experiments comply with our theoretical results concerning the error bounds. We also observe that, while the gain from Phase II gets considerably smaller using the current scheme, the Phase II values seldom do (see Figure 3.10). On a number of instances, the approximation error remains such that it would be unsafe to conjecture that a global optimum has been reached.

3.8 Additional experiments

The algorithmic schemes that we have proposed in this paper lend themselves to several combinations that cannot all be reported in depth. In this section, we focus on the most interesting ones.

no.	1	2	3	4	5
A-5-1	6680.98	6680.98	**6898.14**	**6898.14**	**6898.14**
B-5-1	137.60	**210.38**	**210.38**	**210.38**	**210.38**
A-10-1	9803.50	**9858.76**	**9858.76**	**9858.76**	**9858.76**
A-10-2	**14194.90**	**14194.90**	**14194.90**	**14194.90**	**14194.90**
B-10-1	27.29	70.30	168.22	**168.22**	168.22
B-10-2	88.24	285.91	285.91	285.57	289.49

Table 3.11: Piecewise linear based heuristic (1-5 pieces): second phase solution values.

no.	1	2	3	4	5
A-5-1	169.08	3.24	1.26	0.79	0.17
B-5-1	107.52	97.02	38.16	11.54	7.75
A-10-1	30.68	9.91	0.41	0.78	0.33
A-10-2	45.67	19.72	4.76	0.90	1.54
B-10-1	100.80	912.43	53.76	24.77	18.51
B-10-2	103.62	7.48	6.46	0.94	4.14

Table 3.12: Piecewise linear based heuristic (1-5 pieces): improvement brought about by the second phase (percentage).

no.	1	2	3	4	5
A-5-1	0.01	0.16	0.35	0.70	0.90
B-5-1	0.00	0.13	0.16	0.34	0.64
A-10-1	0.09	1588.94	14.83*	8.85*	12.68*
A-10-2	0.04	5.98	113.08	112.94	623.79
B-10-1	0.04	0.81	1.11	2.73	2.36
B-10-2	0.10	2.16	4.81	2.29	70.78

Table 3.13: Piecewise linear based heuristic (1-5 pieces): CPU times. Quantities with a '*' represent optimality GAPs after the time limit was reached (7200 seconds).

no.	time	iter	model	exact	+local	gain	error
A-5-1	82.14	6	6753.47	5625.32	5803.71	3.17	1128.15
B-5-1	18.20	4	113.57	106.05	210.32	98.32	7.52
A-10-1	2072.45	13	9539.3	9847.80	9858.76	0.11	−308.50
A-10-2	800.80	20	11727.6	11636.90	12235.20	5.14	90.70
B-10-1	135.55	7	176.113	138.52	163.37	17.94	37.59
B-10-2	188.96	13	59.15	58.23	181.84	212.25	0.92

Table 3.14: Bilevel trust region algorithm (7-step function model solved with the primal-dual heuristic): detailed results.

3.8.1 Impact of the scale parameter

The size of the parameter θ has a significant impact on the nature of the logit revenue function. To assess this impact on the performance of the piecewise constant and piecewise linear schemes, we considered 3-step and 3-piece approximations, respectively applied to instance B-5-48. As expected, the approximation error is smaller for the piecewise linear scheme. This is achieved, however, at the expense of CPU time.

3.8.2 Refining the approximations

In Figure 3.11, we compare the solutions found by the step function and piecewise linear approximation schemes on instance A-5-2, as the parameter N (respectively N_{CON}) of constant (respectively linear) pieces increases. Both schemes agree on the optimum (purple curve), and the latter does not vary with either N or N_{CON} (7056.2). As expected, the gap between the model's value and the corresponding logit revenues is much smaller for the piecewise linear approximation scheme. Yet in both cases the rate of decrease is low, and getting a significant improvement in this regard is far beyond our computational capabilities. However, while the piecewise linear scheme might not lead, given finite computational power, to an accurate approximation of the logit flows, numerical evidence suggests that a good concavity region is reached using only very rough approximations, i.e., a small number of pieces.

3.8.3 A primal-dual heuristic

To sidestep the computational limitations of the approximation schemes involving a large number of linear pieces, we consider the use of the primal-dual heuristic scheme previously introduced in the deterministic case by Brotcorne et al. in [15].

The idea underlying the method is to adapt Gauss-Seidel iterations to a bilinear reformulation of the problem, thus addressing the original mixed integer program through a sequence of structured linear programs. The algorithm does not possess the ascent property, but rather generates a sequence of "good" primal solutions.

The linear programs solved within the heuristic procedure are derived as follows.

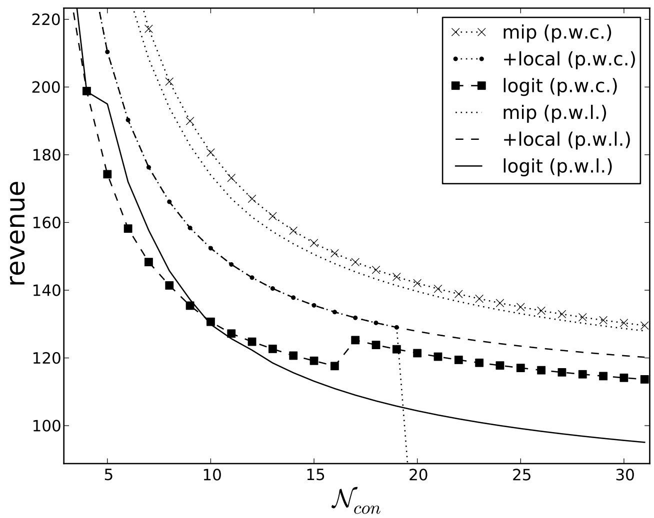


Figure 3.11: Impact of N (resp. N_{con}) on the piecewise constant scheme (resp. piecewise linear scheme) on instance B-5-1: Phase I and Phase II solution values (left); normalized approximation error and percentage gained from Phase II (right).

After replacing the lower level of Program 3.10 (the piecewise linear based formulation) by its necessary and sufficient optimality conditions, which are appended and penalized into the objective, we obtain the linearly constrained program

Program 3.18.

$$\max_{t,w,\pi,\varphi \text{ or } x,w,\pi,\varphi} \quad \sum_{q\in\mathscr{Q}} d_q[t^q \cdot x^q - \tau((c^q+t^q)\cdot x^q + \frac{1}{\theta} 1_{\mathscr{R}^q}\cdot w^q + \frac{1}{\theta}\sum_{n\in\mathscr{N}} \alpha_n\, 1_{\mathscr{R}^q}\cdot \varphi^q_{\cdot n} - \pi_q)]$$

$$\text{s.t.} \quad w - \log\alpha_n x + \alpha_n \geq 0 \qquad n\in\mathscr{N} \tag{3.59}$$

$$-\frac{1}{\theta}\sum_n (\log\alpha_n + 1_{\mathscr{R}})\,\varphi_{\cdot n} + 1^{\mathscr{R}}_{\mathscr{Q}}\pi - c - t \geq 0 \tag{3.60}$$

$$x \in X \tag{3.61}$$

$$\varphi_{\cdot n} \in X \qquad n\in\mathscr{N} \tag{3.62}$$

$$t \in T.$$

In the above, the lower level primal feasibility of Program 3.10 corresponds to constraints (3.59) and (3.61), and the lower level dual feasibility corresponds to (3.60) and (3.62). In the objective, the penalty weight τ is increased from one iteration to the next, gradually enforcing strong duality. The objective still involves a bilinear term which is maximized sequentially with respect to appropriate subsets of variables. In the first phase of any given iteration, maximization is performed with respect to variables x, w, π and φ, while variable t is treated as a constant and set to its previous iteration optimal value. In the second phase, optimization is performed with respect to variables t, w, π and φ, variable x being treated as a constant and set to its previous iteration optimal value. Proceeding this way, each iteration produces two solutions and two objective values. The algorithm is halted once these values agree, up to a predefined tolerance.

Numerical results are shown in Table 3.15 using a 50-step function. As seen in Table 3.15, this strategy fails to improve on previous results. Also, considering column "gain", the first phase solutions are notably inferior to those found by, say, the exact resolution of a 3-step function approximation (Table 3.8). We shall obtain better results with this scheme when combined with the bilevel trust-region algorithm and using larger values of θ, as shown in the next section.

3.8.4 Other combinations

Good results have been obtained using complex combinations, for instance the bilevel trust region warm-started from an optimum of the 3-step function approximation, or the

no.	cpu	model	exact	+local	gain	error
A-5-1	4.60	1657.27	1629.76	6898.14	323.26	0.02
B-5-1	6.14	110.95	106.12	157.50	48.41	0.04
A-10-1	35.88	2515.41	2321.73	8324.20	258.53	0.08
A-10-2	29.61	4142.45	3188.80	10508.50	229.54	0.23
B-10-1	13.23	146.98	138.66	168.22	21.31	0.06
B-10-2	17.53	55.39	53.18	181.88	242.04	0.04

Table 3.15: 50-step function heuristic: primal-dual heuristic solutions details.

embedding of the piecewise linear approximation within the bilevel trust-region framework. Instead of replacing the logarithm function by a first-order approximation, as done in the bilevel trust region algorithm in Section 5, the latter scheme uses a step function where steps are defined adaptively, from one iteration to the next, so as to provide a good approximation around the current iterate. Independently of the strategy used to avoid small flows, let $\hat{x}^k = \mathrm{logit}(t^{k-1})$, assume a 7-step approximation and let the sequence $\alpha^{x\log x,k}$ as

$$\begin{aligned}
\alpha_1^k &= \delta_{\min}, \\
\alpha_2^k &= \hat{x}^k/3 + 2\delta_{\min}/3, \\
\alpha_3^k &= 2\hat{x}^k/3 + \delta_{\min}/3, \\
\alpha_4^k &= \max\{\hat{x}^k, \delta_{\min}\}, \\
\alpha_5^k &= 2\hat{x}^k/3 + 1/3, \\
\alpha_6^k &= \hat{x}^k/3 + 2/3, \\
\alpha_7^k &= 1.0.
\end{aligned}$$

These are the points at which tangents to $x\log x$ are evaluated at iteration k within a model similar to Program 3.10. This optimization problem is usually too expensive to be solved repeatedly with a “big M” scheme, which motivates its substitution by the primal-dual heuristic.

Some of the best results obtained using $\theta = 0.5$ are summarized in Table 3.16 where we also included the scheme described above under the column “det+tr+pwc7+local”. Additional results are displayed in Tables 3.17–3.18 using $\theta = 5$ and $\theta = 10$, respectively.

The average proportion of the demand assigned to shortest paths (evaluated at the end of Phase II) is 95 per cent for the instances of Table 3.17 and 99 percent for Table 3.18. The number of distinct local optima found by the random search is comparable under any setting and includes the best found solution on the majority of instances. If not, the

no.	rand	det+local	pwc3+local	pwl3+local	tr+local	det+tr-pwc7+loc	pwc3+tr+local
A-5-1	**6898.14**	6680.98	**6898.14**	**6898.14**	5803.71	**6680.98**	**6898.14**
B-5-1	**210.38**	**210.38**	**210.38**	**210.38**	**210.32**	137.60	**210.38**
A-10-1	**9858.76**	9803.50	**9858.76**	**9858.76**	**9858.76**	9803.50	**9858.76**
A-10-2	**14118.90**	**14194.90**	**14194.90**	**14194.90**	12207.80	**14194.90**	**14194.90**
B-10-1	**168.22**	27.29	**168.22**	**168.22**	142.34	27.29	79.06
B-10-2	285.70	88.24	285.91	285.91	292.02	**296.07**	285.91

Table 3.16: Algorithms comparison ($\theta = 0.5$): second phase solution values.

no.	#	rand	det+local	pwc3+local	pwl3+local	tr+local	det+tr-pwc7+loc	pwc3+tr+local
A-5-1	10	7688.15	6608.40	**7764.01**	**7764.01**	7536.88	**7764.01**	6457.99
B-5-1	6	**115.10**	48.07	**115.10**	**115.10**	**115.10**	**115.10**	**115.10**
A-10-1	47	**11302.90**	11141.60	10960.80	**11302.90**	10888.10	**11302.90**	**11302.90**
A-10-2	28	**16453.60**	**16453.60**	**16453.60**	**16453.60**	**16453.60**	**16453.60**	**16453.60**
B-10-1	15	**63.09**	**63.09**	39.46	62.48	62.99	**63.09**	62.99
B-10-2	53	301.50	89.97	275.09	**301.63**	276.21	**301.62**	296.00

Table 3.17: Algorithms comparison ($\theta = 5$): second phase solution values.

no.	#	rand	det+local	pwc3+local	pwl3+local	tr+local	det+tr-pwc7+loc	pwc3+tr+local
A-5-1	6	7794.40	6707.21	7596.72	**7870.32**	7244.19	**7870.32**	**7870.32**
B-5-1	4	**110.96**	46.44	**110.96**	**110.96**	107.09	**110.96**	**110.96**
A-10-1	40	**11454.1**	9803.50	**11454.10**	**11454.10**	11284.00	**11454.10**	10137.00
A-10-2	28	**16660.50**	14194.90	**16660.50**	**16660.50**	**16660.50**	**16660.50**	**16660.50**
B-10-1	12	**63.17**	27.29	**63.17**	**63.17**	62.84	**63.17**	62.84
B-10-2	46	**302.02**	88.24	296.38	**302.02**	245.49	**302.02**	296.38

Table 3.18: Algorithms comparison ($\theta = 10$): second phase solution values.

best solution found by random search is within 4 percent of the best solution found. No other algorithm performs better, except for larger values of θ ($\theta = 5$ or $\theta = 10$) where the scheme *det+tr-pwc7+loc* finds all the best solutions.

The step function and piecewise linear schemes agree with the random search on most instances, giving a sense that quasi-optimal solutions have indeed been found. The latter also significantly outperforms the deterministic heuristic on most B-instances.

3.9 Concluding remarks

In this paper, we have conducted a comprehensive algorithmic study of a logit-based pricing problem involving both combinatorial and nonlinear features. Our main conclusion is that the problem can be solved for a near-optimal solution by a combination of mixed integer approximations, which take care of its combinatorial nature, and local ascent methods, which fine-tune the solution obtained by the MIP. The instances involving a small variance proved, as expected, more difficult, and were best addressed by a bilevel trust-region approach where, at each iteration, the trust-region "model" was solved by a primal-dual heuristic.

Two general conclusions can be drawn from our numerical study. First, the smoothing effect introduced by the probabilistic assignment brings a significant reduction of the number of local optima with respect to the deterministic model. This explains the success of randomized local searches in finding the best solutions on most instances. Experimenting with a range of values of the logit scale parameter θ, we observed that θ provides a rather harsh control over the revenue function relative smoothness: either the revenue function contains few optima and a local method is efficient, or nearly all the demand is assigned to shortest paths alone and a local method fails to converge.

Our second conclusion is that simple combinatorial approximations allow to capture the problem's structure, under any settings. Finding instances for which the deterministic heuristic (involving the simplest of models) fails to find the best solution required some crafting on our part (type B instances). While it is true that very elaborate models yield better approximations of the logit flow, this comes at a significant computational cost and rarely impact the concavity region within which the solution lays in the toll space.

Following on the steps of the most successful approaches, we intend to apply them to the mixed logit environment, which is closer to the situation that arises in revenue management. Finally let us mention, on the theoretical side, a challenging and important issue that consists in analyzing the behaviour, as variance increases, of the logit revenue function, and the pace at which it "loses" local optima to eventually turn pseudo-concave, and thus amenable to local ascent.

3.10 Notation

$\mathscr{G} = (\mathscr{A}, \mathscr{N})$	network graph
$\mathscr{A}_{\text{toll}} \subset \mathscr{A}$	toll arcs ($A_{\text{toll}} = \lvert\mathscr{A}_{\text{toll}}\rvert$)
$\mathscr{Q}$	origin-destination pairs ($Q = \lvert\mathscr{Q}\rvert$)
$\mathscr{R}$	paths ($R = \lvert\mathscr{R}\rvert$)
$\mathscr{R}^q \subset \mathscr{R}$	paths for OD pair $q \in \mathscr{Q}$ ($R^q = \lvert\mathscr{R}^q\rvert$)
$\mathscr{R}^q_{\text{toll}} \subset \mathscr{R}^q$	toll paths for OD pair $q \in \mathscr{Q}$ ($R^q_{\text{toll}} = \lvert\mathscr{R}^q_{\text{toll}}\rvert$)
$q(r) \in \mathscr{Q}$	origin-destination pair associated with path $r \in \mathscr{R}$
$d \in \mathbb{R}^{+Q}$	origin-destation demands
$c \in \mathbb{R}^R$	path fixed costs
$c^q \in \mathbb{R}^{R^q}$	restriction of c to the indices associated with paths in $\mathscr{R}_q$, for $q \in \mathscr{Q}$
$c^A \in \mathbb{R}^A$	arc fixed costs
$t \in \mathbb{R}^R$	path tolls
$t^A \in \mathbb{R}^{A_{\text{toll}}}$	arc tolls
$u(t) \in \mathbb{R}^R$	path expected disutilities under tolls t
VAR	variational inequality
SOL	solution set of a variational inequality
$1^{\mathscr{R}}_{\mathscr{Q}}$	$d \in \mathbb{R}^Q \Rightarrow d' = 1^{\mathscr{R}}_{\mathscr{Q}} d \in \mathbb{R}^R : d'_r = d_{q(r)}$
$1_{\mathscr{Q}}$	$1_{\mathscr{Q}} = [1, \ldots, 1]^T \in \mathbb{R}^Q$
$\text{COV}(t) \in \mathbb{R}^{A_{\text{toll}}} \times \mathbb{R}^{A_{\text{toll}}}$	logit toll arc flow covariance matrix under tolls t
$\text{logit}(t) \in (0,1)^R$	logit path probabilities under tolls t
$\theta \in \mathbb{R}^{+R}$	logit scale parameter
$\delta_{\min} \in \mathbb{R}^{+R}$	lower bound on the logit probabilities

$f^{\text{logit}}(t)$	logit revenue under toll policy t
$f^{\text{det}}(t,x)$	deterministic revenue under toll policy t and assignment x
$f(t,x)$	revenue under toll policy t and assignment x
$g^{\text{det}}(t,x)$	deterministic lower level objective under toll policy t and assignment x
$g(t,x)$	entropy based lower level objective under toll policy t and assignment x
$F(t,x)$	gradient of g with respect to t
T	set of arc feasible path tolls
X	lower level feasible set
X_0	compact subset of X such that $0 \notin X_0$
x	network assignment (path space)
π_q	disutility of the shortest path on OD pair q
$N, N_{\text{obj}}, N_{\text{con}}$	number of segments in a piecewise-constant/linear approximation
$\alpha^{h(x)}$	vector of points at which are evaluated tangents to function h
$\beta^{h(x)}$	non differentiable points of a piecewise-linear curve defined on the basis of $\alpha^{h(x)}$
$m^h(\cdot\|y)$	first order approximation of h around y
$F^0(t,x \mid \alpha)$	step-function approximation of F based on the sequence α
$F^1(t,x \mid \alpha)$	piecewise-linear approximation of F based on the sequence α
(t^k, x^k)	trust-region iterates (tolls and flow) at iteration k
δ_k	trust-region radius at iteration k
η	lower bound on the flow in the bilevel trust-region algorithm
$\hat{x}^k$	point around which the trust-region algorithm evaluates sensitivity information at iteration k

CHAPITRE 4

MIXED-LOGIT NETWORK PRICING

Dans le troisième article, nous généralisons le problème de tarification logit à une population non homogène. La demande est allouée aux routes du réseau, répartie entre plusieurs paires origine-destination, par le biais d'un modèle de choix discret logit mixte (*mixed-logit*, en anglais) dans lequel la sensibilité au prix est aléatoire. L'espérance du revenu généré par la demande s'exprime

$$F(t) = \sum_{r \in \mathscr{R}} \int_0^{\alpha_{\max}} \text{logit}_r(t,\alpha)\, f(\alpha) d\alpha\, t_r,$$

où f est la fonction de densité associée à la sensibilité au prix, c est le vecteur de coûts fixes, t est le vecteur de tarifs et $\text{logit}_r(t,\alpha)$ est la probabilité qu'un usager dont la sensibilité au prix est α choisisse la route r :

$$\text{logit}_r(t,\alpha) = \exp[-\theta \sum_{a \in r}(c_a + \alpha t_a)] / \sum_{\ell \in \mathscr{R}} \exp[-\theta \sum_{a \in \ell}(c_a + \alpha t_a)].$$

La fonction de revenu F n'admet pas d'expression analytique en général et son estimation peut réclamer l'utilisation de simulations. Dans cet article, nous montrons que la résolution exacte d'une approximation combinatoire de F permet souvent d'identifier les meilleures régions de concavité et nous présentons des approximations non linéaires permettant d'implanter des algorithmes de montée efficaces. Nous présentons aussi une interprétation économique des solutions et illustrons la richesse du modèle qui permet d'estimer les contributions au revenu de différents segments de la population.

Deux techniques nous permettent d'obtenir des approximations analytiques de F. La première se base sur une approximation discrète de la fonction de densité f. Pour un ensemble de N classes d'usagers, tel qu'un usager appartient à la nième classe avec probabilité β_n et est caractérisé par une sensibilité au prix de γ_n, nous avons

$$F(t) \approx F^{\mathrm{DS}}(t) = \sum_{r \in \mathscr{R}} \sum_{n=1}^{N} \beta_n \,\mathrm{logit}_r(t, \gamma_n)\, t_r.$$

Notons que l'expression de probabilité associée au choix d'une route ci-dessus correspond au modèle *latent class logit* étudié par Green [41] dans un cadre multi-classes discret. Par ailleurs, le problème de maximiser l'approximation discrète-stochastique F^{DS} s'apparente fortement au problème étudié dans notre deuxième article [40].

La seconde technique se base sur une approximation uniforme par morceaux de la distribution du paramètre de sensibilité au prix et ne fait pas intervenir les probabilités de choix logit directement. Considérons un ensemble de N classes d'usagers tel que les usagers de la nième classe soient caractérisés par une sensibilité au prix uniformément distribuée sur l'intervalle $(\alpha_n, \alpha_{n+1}] \subset \mathrm{Image}(f)$, pour $n = 1, \ldots, N-1$. Soit $(f_n)_{n=1}^{N-1}$, les valeurs associées à la fonction étagée correspondante et servant à approximer la fonction de densité f. Nous obtenons alors l'approximation uniforme-stochastique F^{US} telle que

$$F(t) \approx F^{\mathrm{US}}(t) = \sum_{n=1}^{N-1} (f_n - f_{n-1})\, \tau(t, \alpha_n)$$

où $\tau(t, \alpha_n) = \theta^{-1} \log \sum_{r \in \mathscr{R}} \exp[-\theta \sum_{a \in r} (c_a + \alpha_n t_a)]$ représente l'espérance de la désutilité de la route choisie par un usager de la nième classe.

Deux types d'approximations combinatoires sont obtenues après avoir remplacé les probabilités de choix logit (qui interviennent implicitement ou explicitement dans les approximations analytiques ci-dessus) par une affectation déterministe. Pour la sensibilité au prix discrétisée, l'approximation discrète-déterministe F^{DD} s'apparente aux modèles étudiés dans [51] et est une fonction semi-continue supérieurement. Pour la sensibilité au prix uniforme par morceaux, l'approximation uniforme-déterministe F^{UD} est obtenue en substituant aux fonctions τ la désutilité d'un plus court chemin sous l'affectation déterministe. Nous montrons que F^{UD} est un fonction continue des tarifs.

Les différentes approximations, combinatoires et non linéaires, sont illustrées sur de petits exemples, résolus à la main ou numériquement. Nous remarquons alors que la présence de nombreux maxima dans l'approximation discrète-déterministe F^{DD} est due à sa nature discontinue, et ne témoigne pas d'une combinatoire présente dans le vrai problème. Il est apparent à cet égard que l'approximation uniforme-déterministe F^{UD} offre un avantage.

Des essais numériques ont été effectués avec des réseaux de haute densité et où f est donnée par une mixture de deux gaussiennes. Pour la majorité des instances testées, et indépendamment de la valeur du paramètre d'échelle logit, une recherche locale randomisée parvient à identifier la meilleure solution en une centaine de tentatives ; les différents algorithmes s'accordent par ailleurs sur la solution optimale. La situation est analogue à ce que nous avons observé pour le cas d'une demande homogène (deuxième article) où l'affectation logit venait lisser la fonction objectif et produire un problème d'optimisation comportant moins d'optima que sous une affectation déterministe. Dans le cas d'une modélisation logit mixte, nous observons maintenant que l'effet combiné de l'affectation logit et d'une sensibilité au prix aléatoire produit une fonction de revenu encore mieux conditionnée et ne comportant pas un grand nombre d'optima.

Nous avons retenu dans l'article un petit nombre de réseaux pour lesquels les solutions obtenues par les algorithmes diffèrent et permettent de faire quelques commentaires d'une portée générale. D'abord, la précision des approximations est, en général, bien meilleure que dans le cas d'une demande homogène, où nous avons pourtant utilisé des modèles combinatoires beaucoup plus élaborés (deuxième article). Ensuite, l'approximation uniforme-déterministe F^{UD} est la plus difficile à résoudre mais procure des solutions d'une grande qualité, et ce même si un gap d'intégralité important demeure dans la résolution du problème en nombres entiers associé. À temps de calcul égal, comparativement à l'approximation discrète-déterministe F^{DD}, les solutions de l'approximation uniforme-déterministe F^{UD} se trouvent souvent dans de meilleures régions de concavité.

Certains problèmes de recherche opérationnelle sont tels que la meilleure stratégie de résolution consiste à formuler le modèle le plus près possible de la réalité, quitte à se contenter d'une résolution partielle par la suite. Nous avons montré dans le deuxième article que pour une demande homogène, l'approximation combinatoire la plus simple permettait de capturer la structure du problème et d'identifier les meilleures régions de concavité. Nous constatons maintenant que sous une demande hétérogène, on gagne à utiliser une approximation plus réaliste quant à la distribution des classes d'usagers.

Submitted to Computational Optimization and Applications on January 18, 2012.

MIXED-LOGIT NETWORK PRICING

FRANÇOIS GILBERT, PATRICE MARCOTTE AND GILLES SAVARD

ABSTRACT

In this paper, we address a network pricing problem where users are assigned to the paths of a transportation network according to a mixed logit model, i.e., price sensitivity is not assumed to be uniform throughout the user population. We propose algorithms based on combinatorial approximations and show that the smoothing effect induced by both the discrete choice and price sensitivity features of the model actually reduces the number of local optima, and makes it easier to obtain a global solution, compared to simpler models where the combinatorics is predominant. Also, we estimate the proportion of revenue raised from the various population segments, an information that can be used for policy purposes.

keywords: Mixed-logit, combinatorial optimization, network pricing.

January 2012

4.1 Introduction

Let us consider a multiclass extension of the logit network pricing problem analyzed in [39, 40], where the sensitivity to price is not uniform throughout the population. This leads to a mixed logit discrete choice model [75], where the assignment of users to the paths of the transportation network depends on the costs along the arcs (fixed costs plus tolls) and the distribution of the price sensitivity parameter. The numerical challenge associated with this problem is that, in contrast with simpler random utility models, no closed form solution is available for the assignment of flows to a given set of alternatives (paths). The aim of this paper is to address this challenge, and to show that the price sensitivity heterogeneity actually alleviates the nonconcavity of the problem and makes it more amenable to an efficient numerical resolution. Our contribution is twofold. First, we adapt and implement the algorithmic schemes proposed in [39, 40], which are based on a sequence of combinatorial approximations and local search techniques. Next we assess the contribution to the overall revenue of the various segments of the population.

While this paper is the first, to our knowledge, to tackle a revenue-maximizing network pricing problem within a mixed logit environment, it is worth mentioning reference [30], where the goal is to enhance the performance of a congested network through the selection of appropriate tolls. In this work, the authors adopt the probit framework, which leads to a stochastic equilibrium problem. Arguably, the assumption of normal random terms along the arcs induces dependencies between the paths of the network and yields a more general representation of travel behaviour. After showing the differentiability of the equilibrium flows, the authors consider a local descent method based on an analytical approximation of the probit probabilities. A second best network design problem is namely put forth, for which a uniform optimal toll policy is found.

The paper is organized as follows. Section 4.2 is devoted to our revenue maximization model, which is contrasted against the deterministic continuous multiclass formulation analyzed in [56]. In Section 4.3, we present the resolution approach together with the involved nonlinear and combinatorial approximations of the original model. Section 4.4 is devoted specifically to the *discrete-stochastic* approximation and to its combinatorial counterpart, Section 4.5 to the *uniform-stochastic* approximation and to its combinatorial counterpart. In Sections 4.6 and 4.7, we work out in detail two examples involving a single and multiple origin-destination pairs, respectively. Section 4.8 presents numerical results on non trivial instances and, in the concluding Section 4.9, we discuss future developments. Notations are summarized in Section 4.10.

4.2 Problem formulation

The problem we consider consists in maximizing the expected toll raised from commuters assigned to a network according to a mixed logit model. Specifically, we consider a network $\mathscr{G} = (\mathscr{A}, \mathscr{N})$, whose arc set $\mathscr{A}$ is partitioned into a toll set $\mathscr{A}_{\text{toll}}$ and its complement $\mathscr{A} - \mathscr{A}_{\text{toll}}$, and where $\mathscr{N}$ denotes the node set. Let $\mathscr{Q}$ denote the set of OD pairs and, for each $q \in \mathscr{Q}$, let $\mathscr{R}^q$ denote the corresponding path set. Each arc a in the network is endowed with a fixed cost c_a and a toll (to be determined) t_a. The disutility of a path r is modeled by the random variable $\tilde{u}_r$ such that

$$\tilde{u}_r = \sum_{a \in r} (c_a + \tilde{\alpha}\, t_a) + \tilde{\varepsilon}_r,$$

where $\tilde{\varepsilon}_r$ is a zero expectation Gumbel error term, independently distributed in $\mathscr{R}$, with an associate scale parameter $\theta > 0$. The random variable $\tilde{\alpha}$ models commuters price sensitivity and is distributed in the population according to a random variable with density function f. The following notation is also used in the sequel:

$$u_r(t, \alpha) = \mathbb{E}(\tilde{u}_r | t, \tilde{\alpha} = \alpha),$$

which yields $u_r(t, \alpha) = \sum_{a \in r}(c_a + \alpha\, t_a)$ and the associated choice probabilities corresponding to the logit ratios

$$\text{logit}_r(t, \alpha) = \exp[-\theta u_r(t, \alpha)] / \sum_{r' \in \mathscr{R}} \exp[-\theta u_{r'}(t, \alpha)], \quad r \in \mathscr{R}. \tag{4.1}$$

The mixed-logit network pricing problem takes the form

Program 4.1. (continuous-stochastic formulation (mixed-logit))

$$\max_{t} \quad F^{\mathrm{CS}}(t) = \sum_{r \in \mathscr{R}} t_r \, x_r(t) \tag{4.2}$$

$$\text{s.t.} \quad x_r(t) = \int_{\alpha_{\min}}^{\alpha_{\max}} \mathrm{logit}_r(t,\alpha) f(\alpha) d\alpha \qquad r \in \mathscr{R} \tag{4.3}$$

$$\tau(t,\alpha) = -\theta^{-1} \log \sum_{r \in \mathscr{R}} \exp[-\theta u_r(t,\alpha)] \tag{4.4}$$

$$u_r(t,\alpha) = \sum_{a \in r} (c_a + \alpha \, t_a) \qquad r \in \mathscr{R} \tag{4.5}$$

$$\mathrm{logit}_r(t,\alpha) = \exp[-\theta(u_r(t,\alpha) - \tau(t,\alpha))] \quad r \in \mathscr{R}, \tag{4.6}$$

which, for ease of presentation, involves a single OD pair. The expression used to express the logit ratio in equation (4.6) is motivated by our frequent and subsequent use of the quantity τ, the expected disutility of the shortest path. The density vector $x(t)$ corresponds to the mixed-logit assignment obtained in response to the toll policy t.

This formulation can be viewed as a stochastic extension of the deterministic multiclass problem considered in [56]:

Program 4.2. (continuous-deterministic formulation)

$$\max_{t} \quad F^{\mathrm{CD}}(t) = \sum_{r \in \mathscr{R}} t_r \int_{\alpha_{\min}}^{\alpha_{\max}} x_r(t,\alpha) f(\alpha) d\alpha \tag{4.7}$$

$$\text{s.t.} \quad u_r(t,\alpha) = \sum_{a \in r} (c_a + \alpha \, t_a) \qquad r \in \mathscr{R} \tag{4.8}$$

$$0 \leq u(t,\alpha) \cdot x(t,\alpha) - \pi(t,\alpha) \, \mathrm{e} \perp x(t,\alpha) \geq 0 \tag{4.9}$$

$$\sum_{r \in \mathscr{R}} x_r(t,\alpha) = 1, \tag{4.10}$$

where the disutility is expressed in terms of delay rather than cost, where the notation $\perp$ in equation (4.9) is a convenient shorthand to express the complementarity

$$x(t,\alpha)_r > 0 \Rightarrow (c_r + \alpha \, t_r) \cdot x(t,\alpha)_r = \pi(t,\alpha),$$

and where the dual variable $\pi(t,\alpha)$ is the disutility of a shortest path for a commuter with price sensitivity α, under toll policy t. In the deterministic case, price sensitivity $\tilde{\alpha}$ can equivalently be replaced by an appropriate random variable representing the value of time of a commuter and weighting each arc fixed cost, instead of the tolls, as is done in [56]. This equivalence however does not hold in the stochastic case.

4.3 Resolution approaches

While the mixed logit model, under weak hypothesis, corresponds to a differentiable optimization program, it does not lend itself easily to even the simplest ascent methods, since simulation is required to estimate its objective function. Accordingly, our resolution strategy substitutes to the original problem a tractable approximation, which will be solved for a local optimum by a local search method. In order to guarantee a good solution in the global sense, yet another approximation will be solved for its global solution, which will be used to warm-start the initial, finer, approximation. This strategy produces a price vector at which the true value of the objective is then evaluated.

Actually, we consider two tractable approximations schemes for which the objective function assumes a closed form, and obtained by either:

1. replacing the density f by a finite set of mass points (histogram), yielding the *discrete-stochastic* approximation;

2. or replacing f by a piecewise constant density function, yielding the *uniform-stochastic* approximation.

That the logit revenue function assumes a closed form expression under the first assumption is immediate. Indeed, the resulting assignment corresponds to the latent-class logit choice model considered in [41], and the setting is similar to that of [39, 40]. In contrast, the existence of a closed form expression in the piecewise-uniform case relies on the specific structure of the optimization problem. In both cases the resulting mixed integer programs can be solved for a global optimum by off-the-shelf software.

In our implementation, the warm-start approximations simply neglect the random error terms in path disutilities, i.e., the logit probabilities are replaced by an all-or-nothing assignment to a shortest path. We refer to them as *discrete-deterministic* and *uniform-deterministic*, respectively. Obviously, replacing the logit ratio (4.1) by a feasible solution of the complementarity system (4.9) yields a very coarse approximation of the logit flow. The rationale behind this choice is that, based on experience, we observed that the optimal solution of the analytical model ended up in a promising concavity region [40] of the original model for a large proportion of instances. Within this region, the role of the ascent method is simply to refine and improve the values of flows and prices.

Considering the two differentiable approximations and the two combinatorial approximations, we obtain a total of four two-phase algorithmic combinations. These combinations are analyzed in the next two sections.

4.4 Discrete approximations

Let us consider a discrete approximation of the price sensitivity distribution, where the logit assignment of class specific demand is retained. More precisely, we consider a partition of the price sensitivity parameter range $(\alpha_{\min}, \alpha_{\max}]$ into subintervals $(\alpha_n, \alpha_{n+1}]$, for $1 \leq n \leq N-1$, as well as a vector $\gamma \in \mathbb{R}^N$ of price sensitivities such that $\gamma_n \in (\alpha_n, \alpha_{n+1}]$. The elements of the mass probability vector β are then set to the mean value

$$\beta_n \quad = \quad \int_{\alpha_n}^{\alpha_{n+1}} \alpha\, f(\alpha) d\alpha. \tag{4.11}$$

A commuter belonging to class n with probability β_n is characterized by its price sensitivity γ_n. In this setting, the mixed logit pricing problem simplifies to

Program 4.3. (discrete-stochastic approximation)

$$\max_{t} \quad F^{DS}(t) = \sum_{r\in\mathscr{R}}\sum_{n\in\mathscr{N}} \beta_n x_r^n(t) \sum_{a\in r} t_a$$

$$\text{s.t.} \quad u_r^n(t) = \sum_{a\in r}(c_a + \gamma_n t_a) \qquad r \in \mathscr{R}, \quad n \in \mathscr{N} \tag{4.12}$$

$$\tau_n(t) = -\theta^{-1} \log \sum_{r\in\mathscr{R}} \exp[-\theta u_r^n(t)] \qquad n \in \mathscr{N} \tag{4.13}$$

$$x_r^n(t) = \exp[-\theta(u_r^n(t) - \tau_n(t))] \qquad r \in \mathscr{R} \quad , \quad n \in \mathscr{N} \tag{4.14}$$

where, with reference to Program 4.1, $x^n(t) = \text{logit}(t, \gamma_n)$.

The above program is basically an unconstrained program for which a local search method yielding a first-order point can be implemented. It constitutes a generalization of the single-class problem considered and analyzed in [39].

For algorithmic purposes, it is useful to obtain a workable expression for the gradient of the objective function of Program 4.3. The following result is given without proof. Let $p(t,\gamma)$ denote the toll arc choice probability vector, with components

$$p_a(t,\gamma) = \sum_{r:a\in r} \text{logit}_r(t,\alpha). \tag{4.15}$$

Then, the entries of the variance-covariance matrix $\text{COV}(t,\gamma)$ of toll arc choice probabilities satisfy

$$\text{COV}_{ab}(t,\gamma) = p_{ab}(t,\gamma) - p_a(t,\gamma)p_b(t,\gamma),$$

where $p_{a,b}(t,\gamma) = \sum_{r\in\mathscr{R}:a,b\in r} \text{logit}_r(t,\gamma)$. From the equality

$$\partial p_a(t,\gamma)/\partial t_b = \gamma\, \text{COV}_{ab}(t,\gamma)$$

we infer that

$$\partial F^{DS}(t)/\partial t_a = p(t) - \theta \sum_{n=1}^{N-1} \beta_n \gamma_n \text{COV}(t,\gamma_n)\, t.$$

A combinatorial approximation is obtained from Program 4.3 as a zero variance deterministic limiting case. With reference to the probability mass points β_n defined by (4.11) and the associated vector of discrete price sensitivities γ_n, this yields the mathematical program

Program 4.4. (discrete-deterministic approximation)

$$\max_{t,x,\pi} \quad F^{\mathrm{DD}}(t,x) = \sum_{n\in\mathcal{N}} \beta_n\, t\cdot x^n/\gamma_n \tag{4.16}$$

$$\text{s.t.} \quad 0 \le u^n(t)\cdot x^n - \pi_n\, e \perp x^n \ge 0 \quad n\in\mathcal{N} \tag{4.17}$$

$$\sum_{r\in\mathcal{R}} x_r^n = 1 \qquad n\in\mathcal{N} \tag{4.18}$$

$$u_r^n(t) = \sum_{a\in r}(c_a + \gamma_n\, t_a) \qquad r\in\mathcal{R}, \quad n\in\mathcal{N} \tag{4.19}$$

where $e = [1,\ldots,1]^T \in \mathbb{R}^{|\mathcal{R}|}$ denotes the vector of all ones, and x^n the deterministic assignment of β_n commuters having price sensitivity γ_n. Since, for each $r\in\mathcal{R}$ and each $n\in\mathcal{N}$, $x_r^n > 0 \Rightarrow (c_r + \gamma_n t_r)\cdot x_r^n = \pi_n$, we can replace the bilinear objective by its equivalent linear expression $t\cdot x^n = (\pi_n - x^n\cdot c)/\gamma_n$. From there on, one can derive a mixed integer program through the linearization of (4.17), based on the "big-*M*" technique, as proposed by Marcotte et al. in [51] for the single-class problem. Branch-and-bound algorithms can then be used to solve such problems, taking advantage of the discrete nature of the assignment x^n, which can only assume a finite number of discrete values. Note that, in contrast with the logit model, the objective of Program 4.4, as a function of t alone, that is

$$F(t) = \max_x \{F^{\mathrm{DD}}(t,x) | \text{s.t. } (4.17), (4.18) \text{ and } (4.19)\},$$

is piecewise affine and lower semicontinuous (see reference [51]).

4.5 Uniform approximation

The uniform-stochastic approximation differs from the discrete-deterministic in that the price sensitivity density f is approximated by a continuous, piecewise uniform function, rather than a histogram. We introduce an increasing sequence α_n, which defines a piecewise constant density over each interval $(\alpha_n, \alpha_{n+1}]$, for $1 \leq n \leq N-1$, together with the associated set of values

$$f_n = (\alpha_{n+1} - \alpha_n)^{-1} \int_{\alpha_n}^{\alpha_{n+1}} v\, f(v) dv. \tag{4.20}$$

The probability that a price sensitivity value lies in the interval $(\alpha_n, \alpha_{n+1}]$ is equal to $(\alpha_{n+1} - \alpha_n) f_n$. The associated pricing problem is then expressed as:

Program 4.5. (uniform-stochastic approximation)

$$\max_t \quad F^{US}(t) = \sum_n (f_{n-1} - f_n)\, \tau_n(t)$$

$$\text{s.t.} \quad u_r^n(t) = \sum_{a \in r} (c_a + \alpha_n t_a) \qquad r \in \mathscr{R}, \quad n \in \mathscr{N} \tag{4.21}$$

$$\tau_n(t) = -\theta^{-1} \log \sum_{r \in \mathscr{R}} \exp[-\theta u_r^n(t)] \qquad n \in \mathscr{N}. \tag{4.22}$$

The objective function F^{US} is a linear combination of the quantities $\tau_n(\cdot)$ defined in constraint (4.22) as the expected disutility of the shortest path for a commuter having price sensitivity α_n, under a discrete multi-class logit assignment. This program involves the same set of constraints as the discrete-stochastic approximation (Program 4.3), with the exception of disutilities that are expressed in terms of α_n rather than γ_n (i.e., $\alpha_{n-1} < \gamma_n < \alpha_n$) and of the logit probabilities that are not explicit.

Given that the price sensitivity distribution is piecewise-uniform, say given by (4.20), one can show the equivalence of Program 4.1 and Program 4.5. Indeed, we have that

$$\begin{aligned}
\frac{\partial \tau(t,\alpha)}{\partial \alpha} &= -\frac{1}{\theta} \frac{\partial}{\partial \alpha} \log \sum_r \exp[-\theta u_r(t,\alpha)] \\
&= \frac{\sum_r \exp[-\theta u_r(t)] \sum_{a \in r} t_a}{\sum_r \exp[-\theta u_r(t,\alpha)]} \\
&= \sum_r \mathrm{logit}_r(t,\alpha) \sum_{a \in r} t_a,
\end{aligned}$$

from which it follows that the mixed logit revenue F^{CS} (Program 4.1) can be expressed as

$$\begin{aligned}
F^{\text{CS}}(t) &= \sum_r \int_{\alpha_1}^{\alpha_N} \text{logit}_r(t,u) \sum_{a\in r} t_a f(u) du \\
&= \sum_r \sum_n f_n \int_{\alpha_n}^{\alpha_{n+1}} \text{logit}_r(t,u) \sum_{a\in r} t_a du \\
&= \sum_n f_n \, \tau(t,u)|_{\alpha_n}^{\alpha_{n+1}} \\
&= F^{\text{US}}(t).
\end{aligned}$$

For algorithmic purposes, it is useful to obtain a workable expression for the gradient of the objective function of Program 4.5:

$$\begin{aligned}
\frac{\partial \tau}{\partial t_a} &= -\frac{1}{\theta}\frac{\partial}{\partial u_r} \log \sum_r \exp[-\theta u_r(t,\alpha)] \frac{\partial u_r}{\partial t_a} \\
&= \frac{\sum_{r\,:\,a\in r} \exp[-\theta(c_r+\alpha t_r)] t_r}{\sum_r \exp[-\theta(c_r+\alpha t_r)]} \alpha \\
&= \alpha p_a(t,\alpha),
\end{aligned}$$

which yields

$$\partial F^{\text{US}}(t)/\partial t_a = \sum_n (f_{n-1}-f_n)\alpha_n p_a^n(t),$$

where $p_a^n(t) = \sum_{r\in\mathscr{R}:a\in r} \exp[-\theta(u^n(t)-\tau_n(t))]$, $u^n(t)$ is defined as in (4.21) and $\tau_n(t)$, defined in (4.22), provides the class specific toll arc choice probabilities.

Program 4.5 admits a zero variance combinatorial limiting case, obtained after replacing each expected minimum disutility $\tau_n(t)$ by the corresponding deterministic quantity π_n, i.e.,

$$\begin{aligned}
\lim_{\theta\to\infty} \tau_n(t) &= \lim_{\theta\to\infty} \theta^{-1} \log \sum_{r\in\mathscr{R}} \exp[-\theta u_r^n(t)] \\
&= \min\{u_r^n(t) | r \in \mathscr{R}\} \\
&= \pi_n,
\end{aligned}$$

where π_n is feasible for the mathematical program

Program 4.6. (uniform-deterministic approximation)

$$\max_{t,x,\pi} \quad F^{\text{UD}}(\pi) = \sum_n (f_{n-1} - f_n)\, \pi_n \tag{4.23}$$

$$\text{s.t.} \quad 0 \leq u^n(t) \cdot x^n - \pi_n\, e \perp x^n \geq 0 \quad n \in \mathcal{N} \tag{4.24}$$

$$\sum_{r \in \mathcal{R}} x_r^n = 1 \quad n \in \mathcal{N} \tag{4.25}$$

$$u_r^n(t) = \sum_{a \in r} (c_a + \alpha_n\, t_a) \quad r \in \mathcal{R}, \quad n \in \mathcal{N}. \tag{4.26}$$

This program is similar to Program 4.4, except for disutilities that are expressed in terms of α_n rather than γ_n, and the objective of Program 4.6, which does not involve flow variables.

Given that the price sensitivity distribution is piecewise uniform, the continuous-deterministic Program 4.2 simplifies to Program 4.6. Indeed, without loss of generality, based on the perturbation argument invoked in [40], we may assume that class specific shortest paths are unique. This allows to write

$$\begin{aligned}
F^{\text{CD}}(t,x) &= \sum_{r \in \mathcal{R}} \int_0^{\alpha_{\max}} x_r(t,\alpha) \sum_{a \in r} t_a\, f(\alpha) d\alpha \\
&= \int_0^{\alpha_{\max}} \sum_{a \in r_{\min}(t,\alpha)} t_a f(\alpha) d\alpha \\
&= \sum_{n=1}^{N-1} f_n \int_{\alpha_n}^{\alpha_{n+1}} \sum_{a \in r_{\min}(t,\alpha)} t_a\, d\alpha \\
&= \sum_{n=1}^{N-1} f_n\, u_{r_{\min}(t,\alpha)}(t,\alpha) \,\Big|_{\alpha_n}^{\alpha_{n+1}} \\
&= \sum_{n=1}^{N-1} (f_{n-1} - f_n)\, \pi_n \\
&= F^{\text{UD}}(\pi).
\end{aligned}$$

□

Note that the objective (4.23) of the uniform-discrete approximation is a continuous piecewise-linear function of the tolls. This property immediately follows from the disutility of the shortest path π being continuously parameterized by t. Under the assumption that tolls are distinct on distinct paths, π may fail to be differentiable only for a finite number of tolls, those at which the shortest path is not unique. Outside these points, the gradient exists and its components are given by

$$\partial F^{\mathrm{UD}}(t)/\partial t_a = \sum_n f_n[\alpha_{n-1}\, y_a^n - \alpha_n\, y_a^{n+1}],$$

where $y_a^n = \sum_{r\in\mathscr{R}:a\in r} x_r^n$ represents the total flow on toll arc a.

4.6 A single OD-pair example

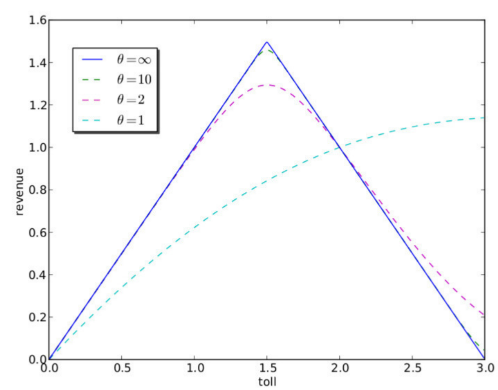


Figure 4.1: Continuous-stochastic models (mixed-logit) for increasing values of θ and the continuous-deterministic model.

In this section, we illustrate the nature of the two discrete-based approximations of the mixed-logit model on a network composed of two parallel links, one of which carries a toll t. In this example, the unique toll path has a zero fixed cost, and the free path has fixed cost s. Commuter price sensitivity is assumed to be uniformly distributed over the segment (a,b).

Figure 4.1 shows the actual revenue associated with four mixed-logit instances, obtained after setting $a = 1/2$, $b = 3/2$, $s = 1$ and a unit demand, for various values of the parameter θ, the limiting case $\theta = \infty$ corresponding to the isosceles triangle. Note that the mixed logit revenue admits a closed form expression, given explicitly as a function of the toll both for bounded values of θ and the deterministic limiting case, which we derive further down. Before considering this case, let us focus on the discrete approximations, as these will help to derive the expression of the mixed logit instances.

In the coarsest approximation, where price sensitivity is equal to one, and setting

$$\text{softmax}_i(a) = \exp(a_i)/\sum_j \exp(a_j)$$

for a given vector a, the discrete-stochastic approximation can be expressed as

$$F^{\text{DS}}(t) = \text{softmax}_1[-\theta(t,s)^T]\,t.$$

This function corresponds to the hatched lines on the left side of Figure 4.2 for distinct values of parameter θ. Setting $\theta = \infty$ (deterministic case) the entire flow is assigned to the toll path for $t < s$, and to the toll-free path for $t > s$. The objective of the discrete-deterministic approximation then takes the form

$$F^{\text{DD}}(t) = H(s-t)\,t,$$

where H is a slightly modified version of the Heaviside step function:

$$H(u) = \begin{cases} 0 & \text{if } u < 0 \\ 1 & \text{if } u \geq 0. \end{cases}$$

The function F^{DD} is plotted as a solid line on the left side of Figure 4.2.

To illustrate the finite multi-class approximation, a uniform discrete framework with

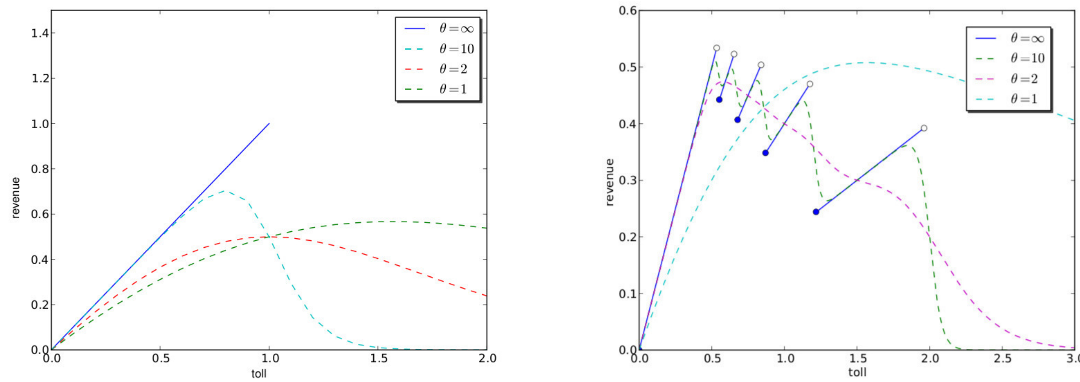


Figure 4.2: Single class (left-hand side) and multiple class (right-hand side) discrete-deterministic and discrete-stochastic approximations.

five equiprobable and equidistant mass points has been adopted:

$$\mathbb{P}(\tilde{\alpha} = \alpha_n) = 1/5,$$

where

$$\alpha_n = a + (b-a)(n-1)/4,$$

for $1 \leq n \leq 5$. The discrete-stochastic approximation takes the form

$$F^{\mathrm{DS}}(t) = N^{-1} \sum_{n=1}^{N} \mathrm{softmax}_1[-\theta(\alpha_n t, s)^T]\, t,$$

shown as the hatched lines on the right-hand side of Figure 4.2, setting $\theta \in \{1, 2, 10\}$. Letting $\theta \to \infty$, the discrete-deterministic approximation can be expressed as

$$F^{\mathrm{DD}}(t) = N^{-1} \sum_{n=1}^{N} H(1 - \alpha_n t)\, t.$$

This corresponds to the discontinuous solid line on the right-hand side of Figure 4.2.

Observe that to each linear segment in the five-class discrete approximation corresponds a subset of commuters perceiving the toll path as the shortest, and discontinuities occur at points where both toll and free-toll paths are shortest. For large values of θ, the five-class discrete-stochastic approximation has several maxima.

As pointed out in Sections 4.3 and 4.5, the mixed logit model admits a closed form expression under a uniform price sensitivity. Under the assumption above, and using the results of Section 4.5, we have

$$F^{\mathrm{CS}}(t) = F^{\mathrm{US}}(t) = \tau(t, 1.5) - \tau(t, 0.5)$$

where $\tau(t,u) = -\theta^{-1}\log[\exp(-\theta\, u t) + \exp(-\theta)]$. This corresponds to the hatched lines of Figure 4.1 for $\theta \in \{1,2,10\}$. As $\theta \to \infty$, the objective of the continuous multi-class deterministic model can be expressed explicitly as a function of the toll:

$$\begin{aligned} F^{\mathrm{CD}}(t) &= F^{\mathrm{UD}}(t) \\ &= \int_a^b H(s - xt)\, t\, f(x)\, dx \\ &= (b-a)^{-1} \int_a^b H(s - xt)\, t\, dx \\ &= -R(s - xt)\,|_a^b\, / (b-a) \\ &= \begin{cases} t & \text{if } t \le s/b \\ (s - at)/(b-a) & \text{if } s/b \le t \le s/a \\ 0 & \text{otherwise}, \end{cases} \end{aligned}$$

where $R(x)$ denotes the Ramp function: $R(x) = H(x)x$. The continuous-deterministic objective $F^{\mathrm{CD}}(t)$ (shown as the solid line in Figure 4.1) is piecewise linear, concave and, unlike its discrete approximation, continuous. The optimal solution is achieved at $t^* = s/b$ for $a > 0$ and at any point of the interval $[s/b, 0)$ otherwise. As $a \to b$, the optimal revenue is left unchanged but the slope of the left-hand side linear segment gets steeper.

Note that the multi-mode nature of approximation F^{DS} (for the larger values of θ), comes from the discretization scheme, and is not related to the combinatorial nature of the pricing problem.

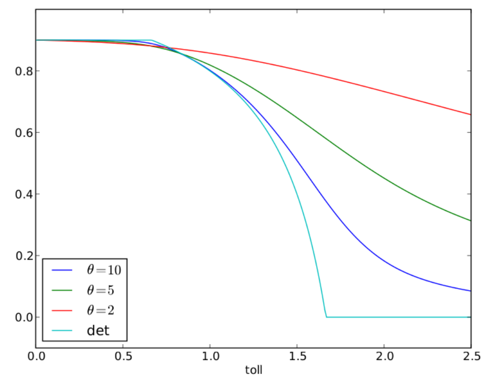


Figure 4.3: Proportion of revenue generated by price sensitive commuters given by the uniform based deterministic and logit approximations.

From the economic point of view, it is interesting to characterize which classes contribute the most to revenue. To fix ideas, we categorize customers as "price sensitive" if their parameter α is above some threshold value c, and "price-insensitive" otherwise. Assume $\tilde{\alpha} \sim U(a,b)$, for some $0 < a < c < b$. The proportion of revenue generated by price sensitive commuters is given by

$$\frac{\tau(t,u)|_c^b}{\tau(t,u)|_a^b} = \log\left(\frac{\exp[-\theta(bt-s)]+1}{\exp[-\theta(ct-s)]+1}\right)/\log\left(\frac{\exp[-\theta(bt-s)]+1}{\exp[-\theta(at-s)]+1}\right)$$

Under the continuous-deterministic model ($\theta = \infty$) this ratio takes the form

$$\begin{aligned}\frac{\int_c^b H(s-xt)t\,dx}{\int_a^b H(s-xt)t\,dx} &= \frac{R(s-xt)|_c^b}{R(s-xt)|_a^b} \\ &= \begin{cases} (b-c)/(b-a) & \text{if } t \le s/b \\ (s-ct)/(s-at) & \text{if } s/b \le t \le s/c \\ 0 & \text{if } s/a \le t \end{cases}\end{aligned}$$

These curves are shown in Figure 4.3 for $a = 1/2, c = 3/5$ and $b = 3/2$. For these

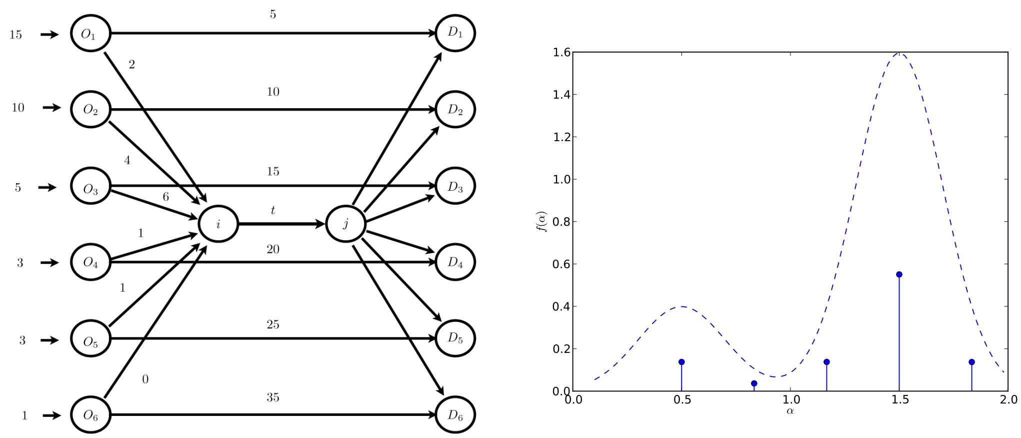


Figure 4.4: Single toll network with multiple OD pairs (left). Bimodal distribution with its discrete and piecewise-uniform approximations.

parameters, price sensitive commuters form 90 per cent of the population.

Whenever the toll is less than or equal to the optimum value $t^* = s/b$, the proportion of revenue generated by price sensitive commuters is equal to their share in the population. In this sense, the deterministic optimum can be perceived as “fair”.

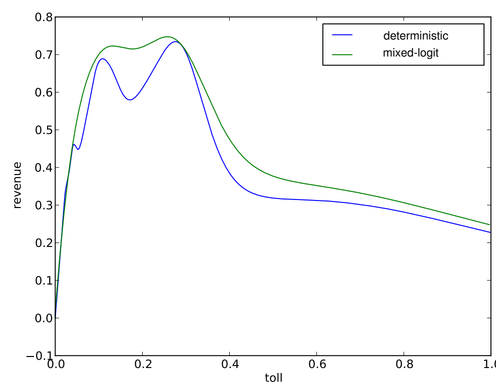


Figure 4.5: Mixed-logit and continuous-deterministic revenues.

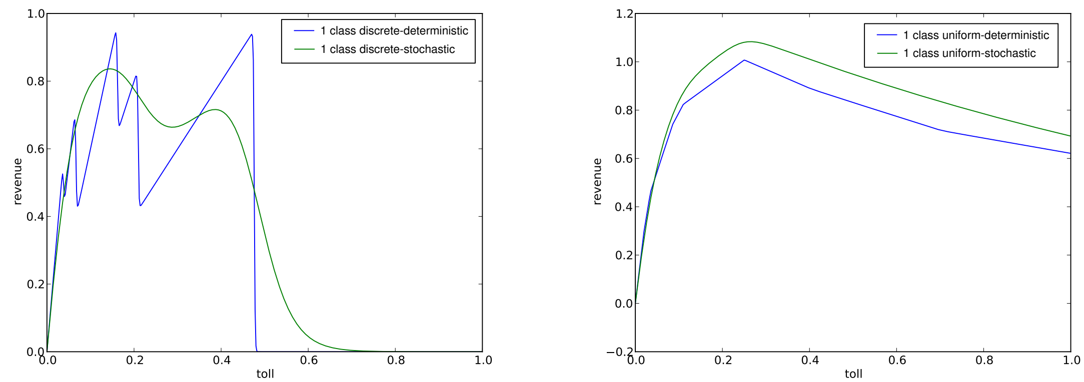


Figure 4.6: One class discrete versus uniform approximation scheme.

4.7 A multiple OD-pair example

In this section, we assess the relative quality of the approximations, both nonlinear and combinatorial, and in the latter case illustrate how the finer uniform-deterministic approximation outperforms the discrete-deterministic approximation. To this aim, we consider the network on the left-hand side of Figure 4.4, involving a single toll arc and six OD pairs, as well as the bimodal price sensitivity density function

$$f(u) \sim 0.2\,\Phi(u|0.5, 0.2) + 0.8\,\Phi(u|1.5, 0.2). \tag{4.27}$$

The graphs of Figure 4.5 illustrate the mixed-logit revenue (continuous-stochastic model) corresponding to $\theta = 0.3$ (green color) and the continuous-deterministic model (blue color).

Figures 4.6 to 4.8 illustrate the discrete approximations (left-hand side) and the uniform approximations (right-hand side), for an increasing number of discrete/uniform classes. We observe that the uniform-deterministic approximations exhibit more robustness with respect to the number of classes. Actually, the uniform-deterministic approximation captures the main features of the mixed model with only a small number of classes and outperforms the discrete-deterministic approximation. It is clear that given

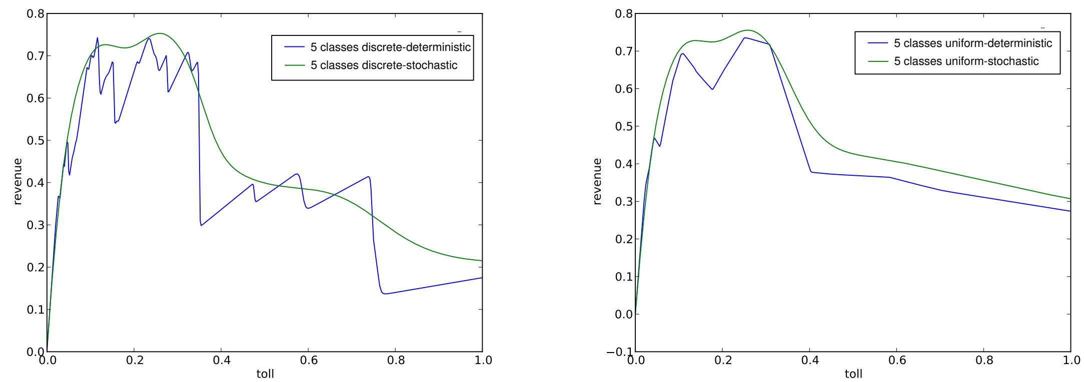


Figure 4.7: Five class discrete versus uniform approximation scheme.

an equal number of discrete or uniform classes, the uniform-deterministic approximation dominates. However, when randomness is taken into account, the discrepancies between the two approximations get very small.

The impact of θ is assessed from an economical perspective in Figure 4.9 where we consider the contribution to the total revenue generated by different commuter segments, under the optimal toll policy. More precisely, we introduce the contribution distribution density function, which describes the probability that a unit of revenue be generated by a given set of commuters. This is illustrated in Figure 4.9 with mixed-logit instances involving a uniform price sensitivity $\tilde{\alpha} \sim U(0.5, 1.5)$ and the bimodal density (4.27). The objective functions appear as the dashed lines on the left-hand side and the right-hand side, respectively. Optimal tolls are obtained for $\theta \in \{0.5, 10\}$ on the right-hand side and $\theta \in \{0.0025, 1\}$ on the left-hand side. Under these settings the proportions of the demand assigned to a shortest path are respectively 97.23 per cent and 100 per cent on the left-hand side, and 64.86 per cent and 99.57 per cent on the right-hand side. The corresponding contribution distributions computed under the optimal tolls are shown as solid lines.

On the right-hand side, for a value $\theta = 0.0025$ which models nearly indifferent commuters, similar proportions of the demand are assigned to each path on each OD pair.

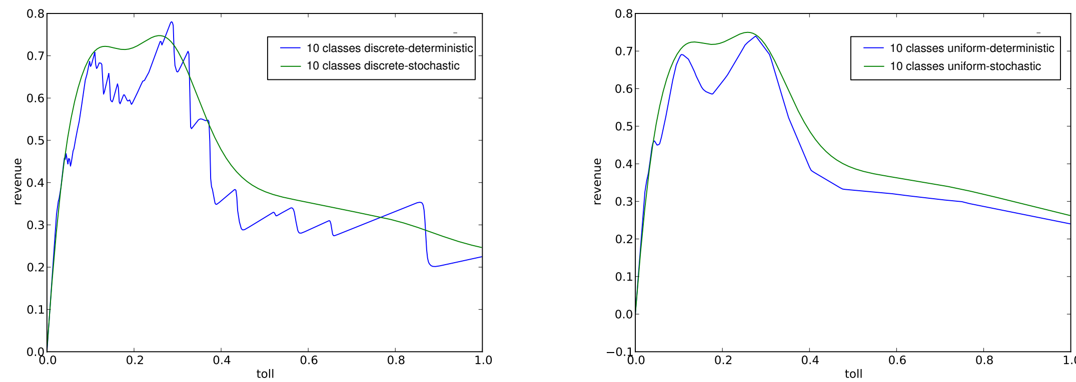


Figure 4.8: Ten class discrete vs uniform approximation scheme.

For such commuters, the contribution distribution is very close to the price sensitivity distribution, and the corresponding optimization problem is easily solved by local search.

The combinatorial features of the problem arise for large values of θ. This can be observed on the left-hand side of the figure, where the rightmost end of each step along the solid blue line for $\theta = 10$ corresponds to the demand associated to an OD pair leaving the more expensive toll path as price sensitivity increases. The discrete features of the contribution density is less obvious for $\theta = 0.5$, even though this yields an increase of less than three percent in demand on non-shortest paths.

These curves can be integrated within the analysis provided at the end of Section 4.6 about the proportion of revenue raised from either price-sensitive versus price-insensitive commuters.

Figure 4.10 shows the revenue function corresponding to the uniform-deterministic approximations for an increasing number of classes (left-hand side), and the mixed-logit models for a range of values of θ (right-hand side) on the network of Figure 4.4. These suggest two numerical approaches. The first was successfully applied to the product line pricing problem by Hanson and Martin [42], a problem involving several demand segments, and in every way similar to ours, minus the network topology. The method

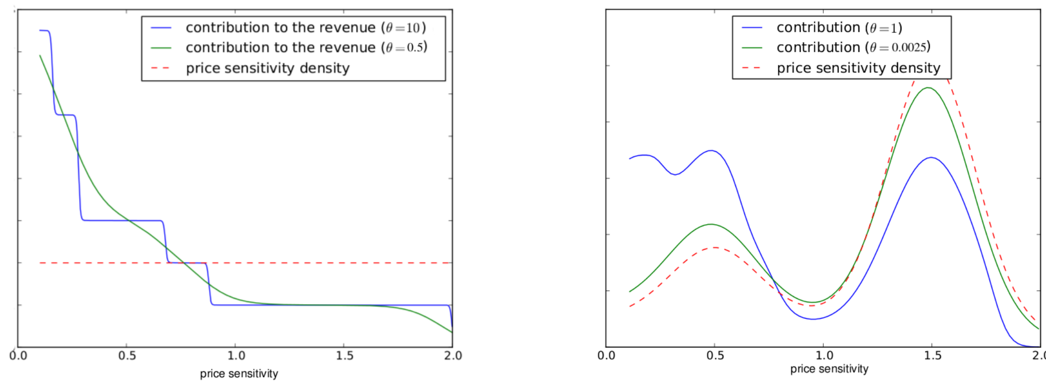


Figure 4.9: Impact of θ on the contribution density function.

consists of a homotopy that employs the logit scale as its parameter. The problem is first solved for a value of θ such that the revenue is unimodal, say $\theta = 0.1$, using a local method. It is then solved iteratively for increasing values of θ, until a target value of θ is reached.

The second resolution scheme, which we adopted, takes a somewhat different perspective. Instead of starting from a high variance model, we solved for its global solution a deterministic approximation, from which a local search was initiated. In other words, a local search on any right-hand side curves is implemented from the global optimum of the corresponding piecewise linear curve on the left-hand side. Figure 4.10 shows that the global optima of most left-hand side models stay within the same concavity region as the mixed-logit right-hand side curves. We argue that this scheme captures the combinatorics induced by the network topology, and allows to determine an optimal or near-optimal solution.

4.8 Numerical experiments

We tested our algorithms on Voronoi, Delaunay or circular topologies (see Figure 4.11). Since a large density graph is required to yield a multimodal optimization problem, we focused on such challenging instances. For instance, the dense circular

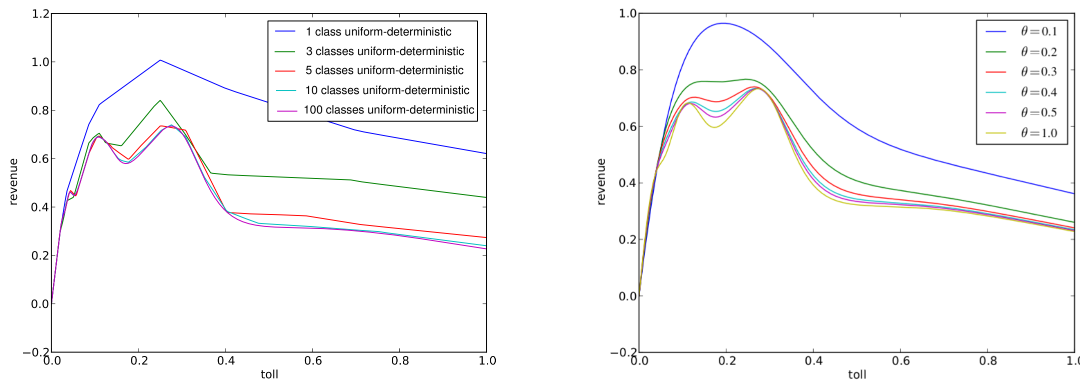


Figure 4.10: Impact of using an increasing number of classes on F^{UD} (left side) and impact of θ on the mixed-logit revenue (right side).

graphs involving sequences of toll and toll-free arcs are not amenable to simplification rules introduced in Bouhtou et al. [12] or Gilbert et al. [39].

While the deterministic formulations are always highly non-concave, this is not the case of the smoothed logit based formulations, no matter how large is the value of the logit scale parameter. Under most settings, and for any given fixed computational time, a random start approach finds the best solution in the vast majority of instances. This behavior has been observed in the case of the single-class pricing problem (see [40]) and is not unexpected, since the multiclass feature compounds the smoothing effect resulting from the logit. This smoothness property actually carries over to any differentiable approximation of the mixed logit model.

The circular network instances involve 20 nodes, 90 arcs, 10 OD pairs and 10 toll arcs. Arc fixed costs are set to zero along the radius and to uniform random numbers between 0 and 100 elsewhere. Ten OD pairs are randomly selected and ten paths of minimal fixed costs, containing one toll-free path, are selected between each OD pair (see [40] for additional details). Price sensitivity follows a Gaussian mixture (see (4.27)). The discrete and uniform approximations are defined according to (4.11) and (4.20) using the appropriate number of classes or uniform segments.

As mentioned above, the approximation schemes put forth in this paper yield four two-phase algorithmic combinations: an ascent algorithm applied to either of the two stochastic approximations (which are differentiable) warm-started from the global optimum of either of the deterministic approximations (which are combinatorial). Since the two ascent algorithms yield almost identical solutions under any setting, we focused on the ascent algorithm applied to the uniform-stochastic approximation, which is less computationally expensive.

Approximation schemes are used at three levels: to obtain a point-wise estimate of the mixed logit revenue, to evaluate the gradient and to define a combinatorial approximation. A 100-class uniform-stochastic approximation is used as point-wise estimate, a 10-class uniform-stochastic approximation is used to evaluate the gradient, and a 5-class approximation to obtain both combinatorial approximations. The local search phase is implemented via the interior point method Ipopt [24], and the combinatorial approximation is solved exactly using Concert CPLEX 10 [1]. Throughout the experiments, the tolerance (integrality gap) is set to 5% in the combinatorial models. CPLEX is halted after 600 seconds, regardless of the gap value.

For small values of θ, such as $\theta = 1$, the problem is always unimodal, and consequently easily solvable by a local search method. For this reason, we set θ to a large value ($\theta = 5$), yet not too large to avoid numerical difficulties in the evaluation of the objective function and its derivatives. In our example, the proportion of the flow assigned to shortest paths, at optimality, was very close to 100 percent.

The various algorithms agreed on the optimum in most experiments. Detailed numerical results are provided for a small set of instances where significant discrepancies could be observed among the solutions. We compare the solutions obtained warm-starting the ascent algorithm from a global optimum of the discrete-deterministic approximation (Table 1) and the uniform-deterministic approximation (Table 2), using a 100 random-start ascent algorithm as a benchmark. Column labels are as follows:

inst	instance number
cpu	time to solve the mixed integer program (mip)
gap	gap
mip	mixed integer program optimal value
exact	uniform-stochastic approximation value under the mip optimal tolls
mip+local	local search from the mip optimum
rand-loc	random-start best solution

Considering the two tables, the quality of the combinatorial approximations, and particularly that of the uniform-deterministic approximation, are far superior to those obtained in the single class case using much more elaborate schemes (see [40]). This is due to the smoothing effect induced by the random price sensitivity.

In the last three instances, the combinatorial approximations are harder to solve, as confirmed by the large gap values achieved after 600 seconds of cpu time. Yet the quality of the final solutions is not correlated to these gaps. This indicates that the right strategy consists in "loosely" solving the uniform approximation, which better fits the initial mixed logit model, independently of the large gaps.

We conclude with the inequalities

$$F^{\mathrm{DS}}(t^1) \leq F^{\mathrm{US}}(t^2) \leq F^{\mathrm{UD}}(t^2) \leq F^{\mathrm{DD}}(t^1) \tag{4.28}$$

where $t^1 \in \arg\max_t F^{\mathrm{DD}}(t)$ and $t^2 \in \arg\max_t F^{\mathrm{UD}}(t)$. While we could not prove their theoretical validity, these relationships were satisfied in all our numerical experiments.

4.9 Conclusion

In this paper, we addressed a generic version of a network pricing problem, where a firm aims at maximizing the revenue raised from tolls set on a subset of arcs of a multicommodity transportation network, and where user perception of travel time (or,

Inst	cpu	gap	mip	logit	+local	rand-loc
1	11	5.00	128.82	91.26	**118.08**	112.38
2	141	5.00	72.40	57.10	61.87	**65.15**
3	18	4.95	83.99	61.18	61.18	**72.17**
4	600	16.03	61.41	52.15	**58.45**	53.52
5	600	18.13	53.25	38.56	**48.84**	44.99
6	600	43.01	114.15	84.18	84.18	**92.06**

Table 4.1: 10-class ascent algorithm warm-started with an optimum of the 5-class discrete-deterministic approximation.

Inst	cpu	gap	mip	logit	+local	rand-loc
1	29	5.00	115.94	113.28	**118.15**	112.38
2	113	5.00	65.38	65.08	**67.07**	65.15
3	44	4.98	75.78	76.50	**80.27**	72.17
4	600	55.81	56.01	58.04	**58.04**	53.52
5	600	90.37	48.14	47.45	**49.60**	44.99
6	600	133.42	95.13	96.52	**105.12**	92.06

Table 4.2: 10-class ascent algorithm warm-started with an optimum of the 5-class uniform-deterministic approximation.

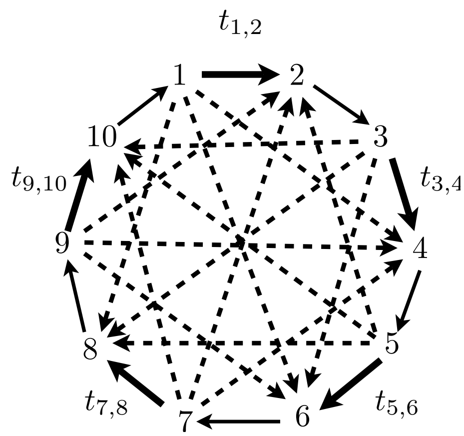


Figure 4.11: 5-toll circular network.

equivalently, price) varies continuously throughout the population. This constitutes a nontrivial extension of the product line design problem, and fits the modern framework of network revenue management (minus the dynamics), an important and challenging research area.

Our contribution is both qualitative and quantitative. From the algorithmic point of view, we implemented approximation schemes that provide starting points from which a local search converges to a near optimal solution, whatever the combinatorial complexity of the problem, i.e., the number of local optima. Next, we found that the combination of two factors, namely the nonuniform behavior of the population, together with the randomness induced by the logit, even in small doses, resulted in smooth and computationally "easy" problems. This provides insights that may be exploited for solving more realistic extensions of the basic problem.

Finally, it is worth mentioning that the model can be adapted to address policy issues, for instance to evaluate the socio-economic impact of private-public partnerships, such as those found in the context of privately managed freeways.

4.10 Notation

θ	logit scale parameter
t	arc toll vector
c	arc fixed cost vector
$\tilde{\alpha}$	price sensitivity (random)
f	price sensitivity density function
$u(t,\alpha)$	path disutility under toll policy t for commuters of price sensitivity α
$u^n(t)$	path disutility under toll policy t in commuter class n
$\tau(t,\alpha)$	shortest path expected disutility under toll policy t for commuters of price sensitivity α
$\tau_n(t)$	shortest path expected disutility under toll policy t for commuter class n
π_n	shortest deterministic disutility for commuter class n
$\text{logit}(t,\alpha)$	logit choice probability under toll policy t for commuters of price sensitivity α
$x(t,\alpha)$	path flow under toll policy t for commuters of price sensitivity α
$x^n(t)$	path flow under toll policy t of commuter class n sensitivity α
$F^{US}(t)$	uniform-stochastic approximation of the mixed-logit revenue function
$F^{DS}(t)$	discrete-stochastic approximation of the mixed-logit revenue function
$F^{UD}(t)$	uniform-deterministic approximation of the mixed-logit revenue function
$F^{DD}(t)$	discrete-deterministic approximation of the mixed-logit revenue function
$F^{CS}(t)$	continuous-stochastic (mixed-logit) revenue function
$F^{CD}(t)$	continuous-deterministic revenue function

CHAPITRE 5

CONCLUSION

Dans cette thèse, nous avons abordé un problème de tarification dans un réseau sous une affectation probabiliste de type logit. Au niveau théorique, nous avons exploité les propriétés analytiques de la distribution logit, dont sa nature markovienne. Nous avons notamment identifié une classe de topologies pour lesquelles le problème est unimodal, et avons à cet égard bon espoir d'élargir ce résultat à une classe de topologies plus grande, par exemple aux réseaux parallèle-sériels.

Plusieurs de nos résultats découlent de la propriété IIA dont témoigne le modèle de choix logit. L'hypothèse sous-jacente de la proportionnalité des élasticités n'a de sens que dans la mesure où les proportions formées par les segments de routes partagées (lorsque comparées deux à deux) demeurent stables. Par ailleurs, les probabilités de choix logit sont fonction de quantités définies sur les routes du réseau uniquement et de telle sorte que l'usager ne voit ni la contribution des différents arcs à la désutilité d'une route, ni la corrélation induite entre les routes par la topologie du réseau. Ceci équivaut à ne pas intégrer la combinatoire induite par cette topologie à la modélisation du comportement. On atteint ici les limites de l'affectation logit et il serait intéressant de généraliser notre analyse à d'autres modèles de choix discret, témoignant de patrons d'élasticité plus élaborés. Le modèle logit emboîté par arc (*link-nested logit*, en anglais), par exemple, fait preuve de réalisme pour la modélisation d'un choix de route et partage, avec le modèle logit, cette propriété markovienne qui nous a permis d'introduire la notion de cellule de réseau.

D'un point de vue numérique, la résolution du problème s'est avérée plus facile que prévu. Nos tests montrent que le terme d'erreur de Gumbel associé à la désutilité d'une route, de même que la sensibilité au prix variable dans le cas d'une demande hétérogène, mènent à une fonction de revenu qui a beaucoup moins d'optima que sous une modélisation déterministe. Bien que nous ayons mis en oeuvre des modèles combinatoires élaborés, ce sont des approximations combinatoires simples qui nous ont permis d'obtenir les meilleures solutions. En complétant la résolution des approximations combinatoires par une recherche locale, nous avons identifié des solutions quasi-optimales pour de nombreuses instances.

Dans le cas d'une demande homogène, nous avons montré à l'aide d'un exemple que l'approximation déterministe peut être, à l'optimum, arbitrairement mauvaise, et ce tant en terme des flots, des tarifs que du revenu. En pratique, cependant, pour la vaste majorité des tests effectués, cette situation ne se présente pas et le modèle de tarification logit et l'approximation combinatoire la plus simple, c'est-à-dire le modèle de tarification déterministe, s'accordent sur la région de concavité contenant un optimum global.

Dans le cas d'une demande hétérogène, aux difficultés relatives à l'optimisation de la fonction objectif est combinée la nécessité d'utiliser des approximations pour évaluer l'objectif. Nous avons d'abord considéré une approximation basée sur une sensibilité au prix discrétisée. L'approximation la plus réaliste du revenu a cependant été obtenue grace à une approximation uniforme par morceaux de la distribution de la sensibilité au prix. Nous avons constaté qu'une résolution même partielle de l'approximation la plus réaliste menait aux meilleures régions de concavité.

Finalement, la méthodologie que nous avons développée peut être adaptée à d'autres programmes bi-niveau non linéaires : un problème de conception de réseau avec congestion ou des contraintes de capacités, une demande élastique ou faisant intervenir un problème d'entropie différent au niveau inférieur.

BIBLIOGRAPHIE

[1] *IBM ILOG CPLEX 10.1.1*. URL `http://www-01.ibm.com/software/integration/optimization/cplex-optimizer/`.

[2] T. Akamatsu. Cyclic flows, markov process and stochastic traffic assignment. *Transportation Research Part B : Methodological*, 30(5):369–386, 1996.

[3] T. Akamatsu. Decomposition of path choice entropy in general transport networks. *Transportation Science*, 31(4):349–362, 1997.

[4] S.P. Anderson et A. De Palma. Multiproduct firms : a nested logit approach. *The Journal of Industrial Economics*, 40(3):261–276, 1992.

[5] G. Aydin et J.K. Ryan. Product line selection and pricing under the multinomial logit choice model. 2000. Working paper, Purdue University, USA.

[6] J.B. Baillon et R. Cominetti. Markovian traffic equilibrium. *Mathematical Programming*, 111(1):33–56, 2008.

[7] E.B. Bajalinov. *Fractional programming : theory, methods and applications*, volume 409. Kluwer Academic Publishers, Dordrecht, 1997.

[8] F. Bastin, C. Cirillo et P.L. Toint. Estimating non-parametric random utility models, with an application to the value of time in heterogeneous populations. *Transportation Science*, 44(4):537–549, 2010.

[9] M. Ben-Akiva et M. Bierlaire. Discrete choice methods and their applications to short term travel decisions. *Handbook of transportation science*, 23:5–34, 2003.

[10] Y. Bentz et D. Merunka. Neural networks and the multinomial logit for brand choice modelling : a hybrid approach. *Journal of Forecasting*, 19(3):177–200, 2000.

[11] M. Bierlaire. The network gev model. Dans *Vortrag zur Swiss Transport Research Conference STRC*, 2002.

[12] M. Bouhtou, S. van Hoesel, A.F. van der Kraaij et J.L. Lutton. Tariff optimization in networks. *INFORMS Journal on Computing*, 19(3):458–469, 2007.

[13] J.J.M. Bront, I. Méndez-Díaz et G. Vulcano. A column generation algorithm for choice-based network revenue management. *Operations Research*, 57(3):769–784, 2009.

[14] L. Brotcorne, F. Cirinei, P. Marcotte et G. Savard. An exact algorithm for the network pricing problem. *Discrete Optimization*, 8(2):246–258, 2011.

[15] L. Brotcorne, M. Labbé, P. Marcotte et G. Savard. A bilevel model for toll optimization on a multicommodity transportation network. *Transportation Science*, 35 (4):345–358, 2001.

[16] J. Cardinal, M. Labbé, S. Langerman et B. Palop. Pricing geometric transportation networks. *International Journal of Computational Geometry & Applications*, 19 (6):507–520, 2009.

[17] E. Cascetta, A. Nuzzolo, F. Russo et A. Vitetta. A modified logit route choice model overcoming path overlapping problems : Specification and some calibration results for interurban networks. Dans *Proceedings of the 13th International Symposium on Transportation and Traffic Theory*, pages 697–711. Lyon, France, Elsevier Science, 1996.

[18] A. Chen, P. Kasikitwiwat et Z. Ji. Solving the overlapping problem in route choice using paired combinatorial logit model. *Transportation Research Record : Journal of the Transportation Research Board*, 1857:65–73, 2007.

[19] K.D. Chen et W.H. Hausman. Mathematical properties of the optimal product line selection problem using choice-based conjoint analysis. *Management Science*, 46 (2):327–332, 2000.

[20] Mei Chen, David H Bernstein et Lazar N Spasovic. Toll-design problem with stochastic route choice. *Environment and Planning B : Planning and Design*, 31 (5):731–742, 2004.

[21] W.C. Chiang, J.C.H. Chen et X. Xu. An overview of research on revenue management : current issues and future research. *International Journal of Revenue Management*, 1(1):97–128, 2007.

[22] S.W. Chiou. Optimization of limited network capacity with toll settings. *Information Sciences*, 179(1-2):109–119, 2009.

[23] C.E. Clark. The greatest of a finite set of random variables. *Operations Research*, pages 145–162, 1961.

[24] *IPOPT 3.3.2*. COIN-OR Project. URL `https://projects.coin-or.org/Ipopt`.

[25] B. Colson, P. Marcotte et G. Savard. Bilevel programming : A survey. *4OR : A Quarterly Journal of Operations Research*, 3(2):87–107, 2005.

[26] B. Colson, P. Marcotte et G. Savard. A trust-region method for nonlinear bilevel programming : algorithm and computational experience. *Computational Optimization and Applications*, 30(3):211–227, 2005.

[27] B. Colson, P. Marcotte et G. Savard. An overview of bilevel optimization. *Annals of Operations Research*, 153(1):235–256, 2007.

[28] R. Cominetti et C. Guzman. Network congestion control with markovian multipath routing. arXiv.org :1107.2900, 2011.

[29] Andrew R. Conn, Nicholas I. M. Gould et Philippe L. Toint. *Trust-region methods*. Society for Industrial and Applied Mathematics, Philadelphia, USA, 2000.

[30] R.D. Connors, A. Sumalee et D.P. Watling. Sensitivity analysis of the variable demand probit stochastic user equilibrium with multiple user-classes. *Transportation Research Part B : Methodological*, 41(6):593–615, 2007.

[31] J.S. Cramer. The origins and development of the logit model. *Logit Models from Economics and Other Fields*, pages 149–158, 2003.

[32] J.M. Day et MA Venkataramanan. Profitability in product line pricing and composition with manufacturing commonalities. *European Journal of Operational Research*, 175(3):1782–1797, 2006.

[33] S. Dewez. *On the toll setting problem*. Thèse de doctorat, Université libre de Bruxelles, Institut de statistique et de recherche opérationnelle, Belgique, 2004.

[34] R.B. Dial. A probabilistic multipath traffic assignment algorithm which obviates path enumeration. *Transportation Research*, 5(2):83–111, 1971.

[35] L. Dimitriou et T. Tsekeris. Elastic multi-user stochastic equilibrium toll design with direct search meta-heuristics. Stuttgart, Germany, 2011. Geiger M.J. and Habericht W.

[36] J.P. Dussault, P. Marcotte, S. Roch et G. Savard. A smoothing heuristic for a bilevel pricing problem. *European Journal of Operational Research*, 174(3):1396–1413, 2006.

[37] S.B. Erlander. *Cost-minimizing choice behavior in transportation planning : a theoretical framework for logit models*. Springer-Verlag, Berlin, 2010.

[38] C. Fisk. Some developments in equilibrium traffic assignment. *Transportation Research Part B : Methodological*, 14(3):243–255, 1980.

[39] F. Gilbert, P. Marcotte et G. Savard. Logit network pricing. Working paper, Université de Montréal, 2012.

[40] F. Gilbert, P. Marcotte et G. Savard. A numerical study of the logit network pricing problem. Working paper, Université de Montréal, 2012.

[41] W.H. Greene et D.A. Hensher. A latent class model for discrete choice analysis : contrasts with mixed logit. *Transportation Research Part B : Methodological*, 37 (8):681–698, 2003.

[42] W. Hanson et K. Martin. Optimizing multinomial logit profit functions. *Management Science*, 42(7):992–1003, 1996.

[43] G. Heilporn, M. Labbé, P. Marcotte et G. Savard. A parallel between two classes of pricing problems in transportation and marketing. *Journal of Revenue and Pricing Management*, 1-2(9):110–125, 2010.

[44] D.A. Hensher et T.T. Ton. A comparison of the predictive potential of artificial neural networks and nested logit models for commuter mode choice. *Transportation Research Part E : Logistics and Transportation Review*, 36(3):155–172, 2000.

[45] W.J. Hopp et X. Xu. Product line selection and pricing with modularity in design. *Manufacturing and Service Operations Management*, 7(3):172–187, 2005.

[46] H.J. Huang et Z.C. Li. A multiclass, multicriteria logit-based traffic equilibrium assignment model under atis. *European Journal of Operational Research*, 176(3): 1464–1477, 2007.

[47] G. Karakostas. Edge pricing of multicommodity networks for heterogeneous selfish users. Dans *In Proceedings of the 45th Annual Symposium on Foundations of Computer Science*, pages 268–276. Press, 2004.

[48] F.S. Koppelman et V. Sethi. Closed-form discrete-choice models. *Handbook of Transport Modelling*, 1, 2000.

[49] F.S. Koppelman et C.H. Wen. The paired combinatorial logit model : properties, estimation and application. *Transportation Research Part B : Methodological*, 34 (2), 2000.

[50] U.G. Kraus et C.A. Yano. Product line selection and pricing under a share-of-surplus choice model. *European Journal of Operational Research*, 150(3):653–671, 2003.

[51] M. Labbé, P. Marcotte et G. Savard. A bilevel model of taxation and its application to optimal highway pricing. *Transportation Science*, 44:1608–1622, 1998.

[52] H. Li et T. Huh. Pricing multiple products with the multinomial logit and nested logit models : Concavity and implications. *MSOM Annual Conference*, 2011.

[53] R.D. Luce. *Individual choice behavior.* John Wiley, Oxford, England, 1959.

[54] B. Maddah et E.K. Bish. Joint pricing, assortment, and inventory decisions for a retailer's product line. *Naval Research Logistics (NRL)*, 54(3):315–330, 2007.

[55] M. Maher. Algorithms for logit-based stochastic user equilibrium assignment. *Transportation Research Part B : Methodological*, 32(8):539–549, 1998.

[56] P. Marcotte, G. Savard et A. Schoeb. Working paper. Université de Montréal, Canada, 2012.

[57] P. Marcotte, G. Savard et DL Zhu. A trust region algorithm for nonlinear bilevel programming. *Operations Research Letters*, 29(4):171–179, 2001.

[58] D. McFadden. *Conditional logit analysis of qualitative choice behavior*, pages 105–142. Frontiers in econometrics. Academic Press, New York, USA, 1973.

[59] A. Neumaier. Complete search in continuous global optimization and constraint satisfaction. *Acta Numerica*, 13(1), 2004.

[60] A. Neumaier, O. Shcherbina, W. Huyer et T. Vinkó. A comparison of complete global optimization solvers. *Mathematical programming*, 103(2), 2005.

[61] A. Papola. Some developments on the cross-nested logit model. *Transportation Research Part B : Methodological*, 38(9):833–851, 2004.

[62] M. Patriksson et R.T. Rockafellar. Sensitivity analysis of variational inequalities over aggregated polyhedra, with application to traffic equilibria. *Transportation Science*, 37(1):56–68, 2003.

[63] Alain Poirier. Gestion de la congestion d'un centre ville par péages. Mémoire de maîtrise, Université de Montréal, Canada, 2007.

[64] J. Prashker et S. Bekhor. Route choice models used in the stochastic user equilibrium problem : a review. *Transport Reviews*, 24(4):437–463, 2004.

[65] S. Roch, G. Savard et P. Marcotte. An approximation algorithm for stackelberg network pricing. *Networks*, 46(1):57–67, 2005.

[66] P. Rusmevichientong, Z.J.M. Shen et D.B. Shmoys. Dynamic assortment optimization with a multinomial logit choice model and capacity constraint. *Operations Research*, 58(6), 2010.

[67] C. Schön. On the optimal product line selection problem with price discrimination. *Management Science*, 56(5):896–902, 2010.

[68] C. Schön. On the product line selection problem under attraction choice models of consumer behavior. *European Journal of Operational Research*, 206(1):260–264, 2010.

[69] Y. Sheffi et W. Powell. A comparison of stochastic and deterministic traffic assignment over congested networks. *Transportation Research Part B : Methodological*, 15(1), 1981.

[70] R. Shioda, L. Tunçel et B. Hui. Applications of deterministic optimization techniques to some probabilistic choice models for product pricing using reservation prices. *Research Report CORR2007-02, University of Waterloo, Canada*, 2007.

[71] R. Shioda, L. Tunçel et T.G.J. Myklebust. Maximum utility product pricing models and algorithms based on reservation price. *Computational Optimization and Applications*, 48(2):157–198, 2011.

[72] A. Sumalee, R. Connors et D. Watling. An optimal toll design problem with improved behavioural equilibrium model : The case of the probit model. *Mathematical and Computational Models for Congestion Charging*, 101:219–240, 2006.

[73] K. Talluri et G. Van Ryzin. Revenue management under a general discrete choice model of consumer behavior. *Management Science*, 50(1):15–33, 2004.

[74] A. Torn et A. Zilinskas. *Global optimization*. Springer-Verlag New York, New York, NY (USA), 1989.

[75] K. Train. *Discrete choice methods with simulation*. Cambridge University Press, Cambridge, Unided Kingdom, 2003.

[76] S. Van Hoesel. An overview of stackelberg pricing in networks. *European Journal of Operational Research*, 189(3):1393–1402, 2008.

[77] D. Van Vliet. Selected node-pair analysis in dial's assignment algorithm. *Transportation Research Part B : Methodological*, 15(1):65–68, 1981.

[78] P. Vovsha et S. Bekhor. Link-nested logit model of route choice : overcoming route overlapping problem. *Transportation Research Record : Journal of the Transportation Research Board*, 1645:133–142, 1998.

[79] T. Yai, S. Iwakura et S. Morichi. Multinomial probit with structured covariance for route choice behavior. *Transportation Research Part B : Methodological*, 31(3): 195–207, 1997.

[80] H. Yang. System optimum, stochastic user equilibrium, and optimal link tolls. *Transportation Science*, 33(4):354–360, 1999.

[81] H. Yang, X. Zhang et Q. Meng. Modeling private highways in networks with entry-exit based toll charges. *Transportation Research Part B : Methodological*, 38(3): 191–213, 2004.

[82] J.Q. Ying et T. Miyagi. Sensitivity analysis for stochastic user equilibrium network flows – a dual approach. *Transportation Science*, 35(2):124–133, 2001.

[83] D. Zhang et D. Adelman. An approximate dynamic programming approach to network revenue management with customer choice. *Transportation Science*, 43 (3):381–394, 2009.