VARIABLE DEMAND PROBIT-BASED NETWORK DESIGN PROBLEM: IMPLICIT PROGRAMMING APPROACH

Richard D. Connors, Agachai Sumalee, David P. Watling
Institute for Transport Studies, University of Leeds, Leeds, LS2 9JT, UK
rconnors@its.leeds.ac.uk, asumalee@its.leeds.ac.uk, dwatling@its.leeds.ac.uk

Abstract

This paper considers the network design problem (NDP) assuming that the users' responses to the design variable follow the Probit stochastic user's equilibrium (SUE). The Probit SUE assumes the normal distribution of the random error terms of link costs. This overcomes the problem of IIA property of the Logit model which is the key issue in route choice. The Probit SUE is extended to the case of variable demand and the fixed point formulation for the variable demand Probit SUE is proposed. This fixed point is imposed onto the formulation of the NDP resulting in a Mathematical Program with Equilibrium Constraints (MPEC). The paper explains the various useful characteristics of the Probit SUE that eases the way for solving the NDP, including the smoothness of the path flows respect to design variables and the uniqueness of path flows. The NDP is then reformulated as an implicit optimization program that allows the application of many fruitful non-linear optimization algorithms to the problem. This paper adopts the Sequential Quadratic Programming (SQP) approach and defines a method for calculating the Jacobian of the objective function respect to the design variable. The optimization algorithm is tested with a simple five-link network.

Keywords: Network design; Probit assignment; Sensitivity analysis; Optimal toll
Topic Area: C6 Network Design, Optimal Routing and Scheduling

1. Introduction

The Network Design Problem (NDP) has a long history in transport research. In such a problem, a planner—with control over some design variables such as toll levels—is assumed to be aiming to optimise some system level measure of performance, while anticipating the effect of drivers adjusting their travel choices in response to the stimuli of the design variables (e.g. Yang & Bell, 1998, 2001; Brotcorne *et al*, 2001; Patriksson & Rockafellar, 2002). Almost exclusively in the NDP literature, it is assumed that drivers possess perfect information on the changing environment, justifying the use of a fixed or elastic demand Wardrop User Equilibrium (UE) model. In spite of questions that could be raised about the realism of assuming perfect information, the UE model has justified its major role in transport planning through its attractive simplicity when used in isolation (i.e. when the values of the design variables are *given*). This has made it highly amenable to analysis; after more than thirty years, still more efficient algorithms continue to emerge for computing UE.

However, when embedded in a NDP, experience has proven quite the reverse to be true of the UE model, with the resulting non-convex, non-smooth optimisation problem providing many, well-documented computational hurdles. These difficulties manifest themselves, and may be characterised, in a number of alternative ways depending on the formulation and solution strategy adopted. However, it might be argued that the three key (to some extent, inter-related) difficulties are:

Active path set changes. Consider first a simplified network structure in which the only paths are non-overlapping and consist of single links, and that for given design variables the travel cost functions are *monotone* (vector of link travel costs continuous and strictly monotonically increasing in the vector of link flows). In this case, for given design variables, there are unique UE link flows and path flows. Now, as the design variables are altered, the set of active (used) paths in the UE solution will vary. The effect is a non-smooth problem in the neighbourhood of any path set change, and in realistic-sized networks with many possible paths, this will happen frequently. This is an example of a wider phenomenon arising from the use of *complementarity problems* (of which UE may be viewed as an example) as constraints to optimisation problems (Patriksson & Rockafellar, 2002, Luo *et al.* 1996). Thus, natural gradient-based search strategies for the NDP may fail even in reaching local optima.

Non-unique UE path flows. The importance of the changing active path set in the success of local search strategies makes it, in some sense, natural to consider the NDP at a *path flow* level. However, in general network structures, while the set of link flow solutions to the UE model at given design variables is a singleton under the assumption of monotone cost functions (Smith, 1979), it is well known that the UE path flow solutions are typically non-unique. Therefore, path-based solution strategies are commonly faced with an additional hurdle of selecting a single UE path flow solution from a convex set, for example by an arbitrary choice of extreme point (e.g. Tobin & Friesz, 1988) or by an additional model selecting the 'most likely' path flows (e.g. Larsson *et al*, 2001). Still, establishing desirable properties of a sequence of such 'unique' UE path flow solutions, as the design variables are altered, may be extremely problematic.

Multiple optima. Even if the difficulties above could be addressed, which cause problems with determining local optima of the NDP, the problem of determining a global optimum would still remain.

Much of the literature on NDP has naturally focused on the algorithmic viewpoint of dealing with such problems, namely attempting to find alternative formulations, ever more sophisticated solution strategies, or clever heuristics that, to some degree, circumvent these difficulties. Taking a step further back, however, we may look at NDP on a more conceptual level, and ask whether such difficulties do not question the initial premise that UE is a sensible choice of model for representing the performance of the network and the responses of users? Would alternative network models, when embedded in the NDP, change the nature of some of these problems? Could these alternative models be defended in their own right, and/or could they be viewed as methods of approximating the original NDP with a UE model?

In particular, the difficulties arising from active path set changes and non-unique path flows naturally bring to mind the Stochastic User Equilibrium (SUE) model (Sheffi, 1985). For problems with monotone cost functions, then under mild conditions on the choice probability model, SUE is know to give rise to solutions (a) in which *all* paths are active, at least in theory, and (b) which are unique in the *path* flow domain (e.g. Cantarella & Cascetta, 1995). Therefore, it is natural to ask, is solving the NDP with an SUE network model actually *easier* than with a UE? At the same time, one is adopting a model that, from a behavioural perspective, is arguably superior in terms of its representation of the uncertainty and heterogeneity that surely exists in traveller decisions.

While the algorithmic advantage of using an SUE model in NDP has been previously implied (e.g. Davis, 1994; Patriksson & Rockafellar, 2003), the approach adopted has been that of a logit-based SUE. Such a model is well known to possess poor realism for network problems, especially in terms of its neglecting of path overlaps, and so loses some attraction as a model in its own right. An improvement is to adopt a nested logit SUE (e.g.

Gentile & Papola, 2001), but such an approach cannot be applied to arbitrary network structures (for almost all realistic network structures, it will not be possible to define a unique hierarchy for the available routes). While developments in discrete choice theory open up many further possibilities, there remains one long-established model that has been known for many years to deal with such problems, namely the probit SUE, and this will be the approach adopted in the present paper. The NDP with the probit SUE in this paper is re-cast into the form of an implicit program.

The structure of the paper is as follows. In Section 2 the necessary notation is introduced. Fixed demand network assignment models are reviewed in Section 3, then in Section 4 methods for extending these to the variable demand case are considered. Section 5 discusses the Network Design Problem. Section 6 presents numerical results of the NDP with a small problem and conclusions are presented in Section 7.

2. Notation

The network itself is represented by a directed graph consisting of *N* nodes, with $\mathcal{A}$ the set of connecting links. The demand matrix, $\boldsymbol{q}$, has entries, q_{rs}, representing the travel demand from origin *r* to destination *s*, where *r*, *s* = 1…*N*. The vector of link flows is $\mathbf{x}$, with link costs $\mathbf{t}(\mathbf{x})$, so that $t_a(x_a)$ is the cost of travelling along link $a \in \mathcal{A}$ when the link flow is x_a. The set of paths connecting node *r* to node *s* is K_{rs}. The (binary) link-path incidence matrix, Δ^{rs}, with elements $\delta^{rs}_{a,k}$, denotes the links comprising each path connecting node *r* to *s*. An assignment of flows to all paths is denoted by the vector $\mathbf{f}$, with $f^{rs}_k \geq 0 \ \forall k, r, s$ .The assignment $\mathbf{f}$ is *feasible* for demand $\mathbf{q}$ if and only if

$$\sum_{k \in \mathcal{K}_{rs}} f^{rs}_k = q_{rs} \quad \forall r, s, \tag{1}$$

and the (convex) set of feasible path flows is denoted *F*. With $\mathbf{c}(\mathbf{f})$ the vector of path costs, the cost of the *k*-th path is

$$c^{rs}_k(\mathbf{f}) = \sum_{a \in A} t_a(\mathbf{x}(\mathbf{f}))\, \delta^{rs}_{a,k}\,.$$

3. Fixed demand network assignment models

The problem of describing the distribution of traffic through a road network has produced a variety of solutions centred on the notion of an equilibrium set of traffic flows that are in some sense stable; no driver having any motivation to change their (route choice) behaviour when the network flows are at equilibrium. The concept of network equilibrium traffic flows was first defined by Wardrop (1952) whose “User Equilibrium” (UE) is based on the behavioural assumption that drivers choose the route minimising their travel cost. The UE flows are deterministic, contingent upon the number of trips between the nodes of the network and the nature of the link cost functions. One characterisation of user equilibrium is:

At UE, no driver can reduce their travel cost by unilaterally changing route.

Travel cost refers to all factors that the analyst may wish to include, whether this is simply travel time, or a more complete generalised measure of travel cost. This definition implies that, at user equilibrium, for each origin-destination (OD) pair the traffic is distributed so that the travel cost on all used paths is equal, and less than the travel time that would be experienced by a single vehicle on any unused path. Not all of the available paths are necessarily used. The UE condition for a set of network flows ($\mathbf{f} \in F$) is that

$$f_k^{rs} > 0 \Rightarrow c_k^{rs}(\mathbf{f}) \le c_j^{rs}(\mathbf{f}) \quad \forall j,k \in K_{rs}, j \ne k, \forall r,s. \qquad (2)$$

The UE traffic assignment makes strong assumptions; that drivers are identical, perfectly rational and have complete knowledge of the network conditions. It does not allow drivers to act on differing individual assessments of the current network conditions or to have different perceptions of the attractiveness of alternative routes. While the UE flows provide a good starting point from which to understand the behaviour of traffic in the network, a more general approach can be adopted that incorporates the heterogeneity of drivers.

In an attempt to incorporate into this traffic assignment model the fact that individual drivers have their own assessment of both network conditions and of the cost of taking different routes (including their personal preferences for some routes over others), their route choice behaviour can be assumed to follow a random utility model. This assumption leads to a different set of equilibrium flows: the Stochastic User Equilibrium (SUE). Let the perceived cost of the *k*-th route be (the random variable) C_k, then

$$C_k = c_k + \varepsilon_k$$

where $c_k = c_k(\mathbf{f})$ is the mean perceived route cost and the random errors $(\varepsilon_1, \varepsilon_2, ...)$ follow some joint probability density function with zero mean vector. Many error structures have been proposed for SUE, not only the commonly used independent Weibull and multivariate Normal that lead to the Logit and Probit models respectively (Sheffi 1985), but also more general cross-nested logit models (Prashker & Bekhor, 1999), mixed error component models (Nielsen *et al*, 2002) and gamma link component distributions (Cantarella & Binetti, 2002). However, the definition of the SUE assignment does not require us to specify the error distribution, so we can return to this issue later.

We then define the proportion, $P_k^{rs}(\mathbf{c})$, of drivers travelling from *r* to *s* choosing the *k*-th path as those who perceive that this is the cheapest route, given path costs **c**;

$$\begin{aligned} P_k^{rs} &= \Pr\left(C_k^{rs} \le C_j^{rs} \ \forall j \in K_{rs}, j \ne k\right) \\ &= \Pr\left(\varepsilon_k^{rs} + c_k^{rs} \le \varepsilon_j^{rs} + c_j^{rs} \forall j \in K_{rs}, j \ne k\right), \end{aligned} \qquad (3)$$

where Pr(.) denotes probability. For most choices of error distribution, in particular for the multivariate normal distribution that leads to the probit case, calculation of this probability (for every path) is one of the central computational difficulties in determining the SUE flows. One of the main attractions of using independent Weibull-distributed error terms (the Logit model) is that these probabilities are straightforward to evaluate exactly. For most other distributions the path choice probabilities cannot be written in closed form and instead must be approximated analytically or estimated via Monte Carlo simulation.

The Stochastic User Equilibrium (SUE) is defined to be a set of flows such that:

At SUE, no driver can improve their perceived travel cost by unilaterally changing route.

The SUE path flow assignment (for $\mathbf{f} \in F$) is the solution to the following fixed-point problem

$$f_k^{rs} = q_{rs} P_k^{rs}(\mathbf{c}(\mathbf{f})) \quad \forall k \in K_{rs}, \forall r,s. \qquad (4)$$

This states that, for a given OD pair, the flow on the *k*-th path consists of those drivers who perceive this to be the best path. Since **f** is defined to be a feasible set of flows, the total number of drivers on all paths connecting *r* to *s* matches the total travel demand from this origin to this destination. A network path flow vector satisfying SUE will be denoted **f***.

If the random error terms follow a distribution that assigns non-zero probabilities across an unbounded domain, then $P_k^{rs} > 0 \ \forall k,r,s$ and every route in K_{rs} is assigned some flow. This is the case, for example, with the logit and probit models. In practice, two factors ameliorate this unrealistic prediction. Firstly, the OD demand is finite and fractions of this demand that represent less than a single vehicle can be ignored. Secondly, when calculating the flows for a given network, the demand will be shared amongst only those paths that appear in K_{rs} and hence the modeller can prevent even minimal flows being assigned to unrealistic routes by omitting them from the path-set, though this requires careful checks to ensure the validity of the restricted path-set. An alternative that avoids these issues is to choose a bounded probability distribution, though this may sacrifice desirable analytical properties of standard distributions.

4. Variable demand network assignment models

It is unrealistic to assume that no matter how severe congestion becomes, no one will decide to postpone (or cancel) their intended journey. Both of these methods for assigning traffic to a network (UE and SUE) can be extended to accommodate the fact that the number of travellers is not fixed. Typically (summary in Sheffi 1985) this variability has been encapsulated by a demand function that describes how the number of travellers depends on the cost of their desired journey. In addition to assigning traffic in accordance with the network equilibrium, this demand variation constitutes another equilibrium condition that must be satisfied. This approach is described for both UE and SUE below. Then for the SUE case an alternative method is presented.

4.1 UE with variable demand

Variable demand is incorporated into the UE model (see for example Beckmann *et al*. 1956) by writing the OD demand q_{rs} as a function, $D_{rs}(.)$, of travel cost. For a given OD pair, the UE travel cost is the minimum path cost, hence solving UE with variable demand requires path flows $\mathbf{f} \in F$ to be found such that for each OD pair the following conditions simultaneously hold:

$$f_k^{rs} > 0 \Rightarrow c_k^{rs}(\mathbf{f}) \le c_j^{rs}(\mathbf{f}) \quad \forall j,k \in K_{rs}, j \ne k$$

$$q_{rs} = D_{rs}\left(\min_{k \in K_{rs}}(c_k^{rs})\right).$$

Recall that, in UE, the minimum cost is exactly what will be experienced on any used path.

4.2 SUE with variable demand: mean perceived cost

For the SUE case there is no agreement amongst drivers on the cost of any particular used route. This means that the UE approach (above) cannot be used directly since the argument required for the demand function for each OD pair is the route cost appropriate for the population as a whole on that movement.

One way to deal with this (Ben-Akiva *et al*. 1986, Gentile & Papola 2001, Maher *et al*. 1999) is to use the expected minimum perceived route cost as the argument of the demand function. This is perhaps the nearest equivalent quantity to the experienced (minimum) route cost that occurs in the UE case. The SUE flows can then be calculated, as in the UE case, with an additional demand equilibrium condition to be satisfied

$$f_k^{rs} = q_{rs} P_k^{rs}(\mathbf{c}(\mathbf{f})) \quad \forall k \in K_{rs}, \forall r, s$$

$$q_{rs} = D_{rs}\left(E\left[\min_{k \in K_{rs}} (C_k^{rs}) \right] \right) = D_{rs}\left(E\left[\min_{k \in K_{rs}} (c_k^{rs} + \varepsilon_k^{rs}) \right] \right) \; \forall r, s.$$

This definition of variable demand SUE is appropriate to any (bounded, monotonically decreasing) demand function $D_{rs}(.)$. One particular choice for the demand function can be derived from the assumption that the drivers' choice of whether or not to travel follows random utility theory, which has already been used to describe their route choice behaviour in the formulation of SUE. Writing the cost of travelling as $C_T = c_T + \varepsilon_T$ and the cost of not travelling as $C_0 = c_0 + \varepsilon_0$, gives the number travelling (the demand) to be

$$q_{rs} = Q_{rs}\left(1 - P_0^{rs}\right) = Q_{rs}\left(1 - \Pr\left(c_0^{rs} + \varepsilon_0^{rs} \le c_T^{rs} + \varepsilon_T^{rs}\right)\right)$$

where the constant Q_{rs} is the total number of travellers considering whether or not to travel, and $(\varepsilon_T, \varepsilon_0)$ follow some zero mean joint probability distribution.

For UE, the cost of travel from *r* to *s* is simply the minimum route cost for that movement, and so the number choosing to travel is obtained from the above equation by substituting $c_T^{rs} = \min_{k \in K_{rs}} \{c_k^{rs}\}$. For SUE, using the expected perceived minimum route cost gives the travel demand:

$$q_{rs} = Q_{rs}\left(1 - \Pr\left(c_0^{rs} + \varepsilon_0^{rs} \le E\left[\min_{k \in K_{rs}} \{c_k^{rs} + \varepsilon_k^{rs}\} \right] + \varepsilon_T^{rs} \right) \right)$$

The random utility model used here to describe the choice of whether or not to travel can assume any distribution (Weibull, Normal, Gamma) for the random error terms $(\varepsilon_T, \varepsilon_0)$, regardless of that used in the route choice model. The problem with this formulation is that, in deciding whether or not to travel, drivers are described as comparing the cost (disutility) of not travelling with the perceived minimum route cost *averaged across the entire population*. It seems more plausible that each driver compares the cost of not travelling with the minimum route cost *as judged by them alone*, i.e. compares the option of not travelling with the single best route as they see it, regardless of the thoughts of the rest of the population (that are unknown to them). This individual-based approach to the variable demand SUE problem derives entirely from random utility theory and leads to a rather simple statement of the variable demand SUE flows.

4.3 SUE with variable demand: individual perceived cost

Consider the drivers' decision as a choice between the utilities of different routes, measured against the disutility of not travelling (or going later or by a different mode that is outside of the road network). For each OD pair, the option of “no travel” can be represented in the network by a pseudo-link that provides drivers with another choice of OD route. In this way, every driver is assigned to the network, as for the fixed demand case; those choosing not to travel are assigned to the pseudo-link connecting the relevant OD pair. The perceived cost of no travel will vary across a population of drivers and so this cost has a random part associated with it, just like the costs of travelling along the other links in the network. However, since the option of not travelling does not suffer from congestion, it is assigned a constant cost (the mean disutility across the population). Writing the cost of not travelling as $C_0 = c_0 + \varepsilon_0$, the proportion of drivers not travelling is

$$P_0^{rs} = \Pr\left(c_0^{rs} + \varepsilon_0^{rs} \le c_k^{rs} + \varepsilon_k^{rs} \,\forall k \in K_{rs}\right)$$
$$= \Pr\left(c_0^{rs} + \varepsilon_0^{rs} \le \min_{k \in K_{rs}} \left\{ c_k^{rs} + \varepsilon_k^{rs} \right\}\right),$$

exactly like the route choice probabilities in (4). The option of whether or not to travel sits alongside the choice of which route to take, each having a distribution of perceived costs across the population of drivers. The people choosing not to travel are those for whom all routes appear more expensive than the disutility of not visiting their desired destination.

The variable demand SUE formulation can then be written as an extended version of the fixed point definition of SUE

$$f_k^{rs} = q_{rs} P_k^{rs}(\mathbf{c}(\mathbf{f})) \quad \forall k \in K_{rs}^0, \forall r, s, \tag{5}$$

where $K_{rs}^0 = K_{rs} \cup \{0\}$, with f_0^{rs} the number of drivers electing to not travel, and q_{rs} is now the (fixed) total number of potential drivers, some of whom choose route 'zero' and do not travel. This mechanism for incorporating variable demand into the SUE model is a natural extension of the fixed demand case and is based on the same underlying principle of random utility theory. The variable demand SUE flows can be calculated exactly as in the fixed demand case, which itself occurs as the special case where the mean perceived cost of not travelling forces everyone to travel, in the limit $c_0 \to \infty$. This model also subsumes the UE case, which can be recovered by taking the limit $\Sigma \to 0$ for the covariance matrix of the joint probability distribution of the random errors. It remains to solve this fixed-point problem in order to calculate the equilibrium flows for a given network. Fortunately, with this formulation, it is no more difficult than calculating the fixed demand SUE flows, methods for which have been extensively discussed elsewhere (Sheffi 1985, Sheffi & Powell 1981, Fisk 1980).

Thus far no restriction has been made on the distribution of the error terms for the route choice model of SUE or for the demand model concerning the choice of whether or not to travel. If all the error terms involved are assumed to be Weibull distributed, then the resulting logit choice model may be equivalently represented as a nested logit choice, thus recovering the same formulation presented in section 4.2. However, for other error distributions (such as the probit case) the methods are distinct. For route choice, we choose to use the probit model which properly accounts for the correlations between overlapping routes (whereas the standard logit model for example does not). For the demand model, various distributions could be argued as valid; adopting the probit model here is desirable as it eases the inclusion of the demand model as a pseudo-link, which then requires no special treatment in the variable demand SUE calculations.

5. Probit-based network design with variable demand

Network Design refers to the `improvement' of a network achieved by changing the network design parameters. This process requires some means of quantitatively judging the optimality of a given network; in its most general form this evaluation might use an objective function comprising a weighted sum of the network's performance against an array of criteria such as total travel time, revenue generated by tolls and measures of reliability, congestion and pollution. Desirable attributes will appear with positive multipliers in the objective function, whereas undesirable attributes will be weighted negatively. The Network Design Problem (NDP) is then to change the network and/or the values of any controllable policy variables (such as tolls) in order to achieve a “maximum score” as evaluated by this objective function.

The only necessary attribute of the objective function is that it can be evaluated at any state of the network, that is, at any given set of link flows. It is usual that the network is evaluated when it has reached (some sort of) equilibrium; this may be at User Equilibrium (UE), Stochastic User Equilibrium (SUE), or at some other well-defined state that allows comparisons to be made from one set of design variables to the next.

The criteria included in the objective function (and their relative weights) determine how the network needs to be changed in order to maximise its performance. A different objective function should be expected to lead to a different set of optimum network flows.

In any given instance of the NDP, there will be only certain features of the network infrastructure that can be changed (e.g. the traffic capacity of some links) these variables are the “network design parameters”. Changing the set of network design parameters is likely to alter the optimum network performance that can be achieved.

5.1 The network design problem as an MPEC

If we choose to evaluate the performance of the network at some equilibrium condition, say SUE, the Network Design Problem naturally presents itself as a mathematical program with equilibrium constraints (MPEC). The mathematical program is an optimisation problem, adjusting the network design parameters (link capacities, tolls etc) in order to maximise the network’s performance as measured by our objective function, and this is constrained to be evaluated with the network flows at SUE.

With objective function $g(.)$, and network design parameters $\mathbf{s}$, the NDP is

$$\begin{array}{rll} \max\limits_{\mathbf{s}} g(\mathbf{f},\mathbf{s}) & \} = \text{upper level} \\ \text{subject to} \quad \mathbf{f} = \mathbf{f}^*(\mathbf{s}) & \} = \text{lower level} \end{array} \tag{6}$$

Solving the NDP typically involves a numerical search across values of the network design parameters to optimise the upper level objective function, whilst evaluating the lower level equilibrium flows at each iteration. The NDP can present a non-smooth objective function ‘surface’ with multiple optima in a high dimensional space. Sophisticated numerical methods are required to tackle such problems. To anticipate some of the difficulties that can be expected in larger networks, and to understand the influence of the choice of equilibrium condition (UE or SUE), we begin by examining a very simple NDP, but one that can provide multiple optima of the objective function.

5.2 Example: Two link network

Consider the two link network with total demand $Q = 11$, with a single network design parameter representing a toll, τ , imposed on Link 1, and link cost functions

$$\begin{array}{ll} C_0(x_0) = c_0 + \varepsilon_0 & \text{with } \varepsilon_0 \sim N(0,\sigma_0^2) \\ C_1(x_1) = 10 + \tau + x_1 + \varepsilon_1 & \text{with } \varepsilon_1 \sim N(0,\sigma_1^2) \\ C_2(x_2) = 60 + x_2^2 + \varepsilon_2 & \text{with } \varepsilon_2 \sim N(0,\sigma_2^2) \end{array}$$

As described in section 4.4, the demand variation is modelled by a ‘pseudo-link’ with cost C_0, and x_0 denoting the number of drivers not travelling. We explore the behaviour of this network, that is to say the SUE link flows, as demand (via c_0) and toll are varied.

Different values for the variance of the choice probabilities are calculated, including the limiting UE case. Figure 1 depicts the UE and SUE flows on the network for various settings of c_0 and τ. When the cost of not travelling is low, all the potential demand is taken by the pseudo-link and no one travels. As the cost of staying at home, c_0, increases, more people are forced onto the network. At low values of the link 1 toll, drivers choose to travel on link 1, since the free flow cost on link 2 is comparatively high. As link 1 becomes congested through either more travellers (as c_0 increases), or becomes less attractive (as τ increases), so more traffic moves onto link 2. Note that the transition of flow between the links as the network parameters change occurs smoothly for SUE. This may be contrasted with equivalent UE flows that change between routes abruptly as the network parameters are changed. This can be seen as the 'sharp corners' on the surfaces of the UE flows in Figure 1 and the cross sections of these surfaces in Figure 2.

Consider the UE flows. Regardless of the toll level, as the cost of not travelling (c_0) increases, more people elect to travel on the network, that is, on links 1 and 2. The toll determines the relative attractiveness of link 1 or link 2, as can be seen in Figure 2 by the shape of the curves when τ =60,90; for these cases the uncongested cost of link 1 is higher than that on link 2, hence the flow is initially assigned to link 2 and only when the congestion on this link increases does link 1 begin to be used. The same behaviour occurs for the SUE case but the transitions in flow occur smoothly.

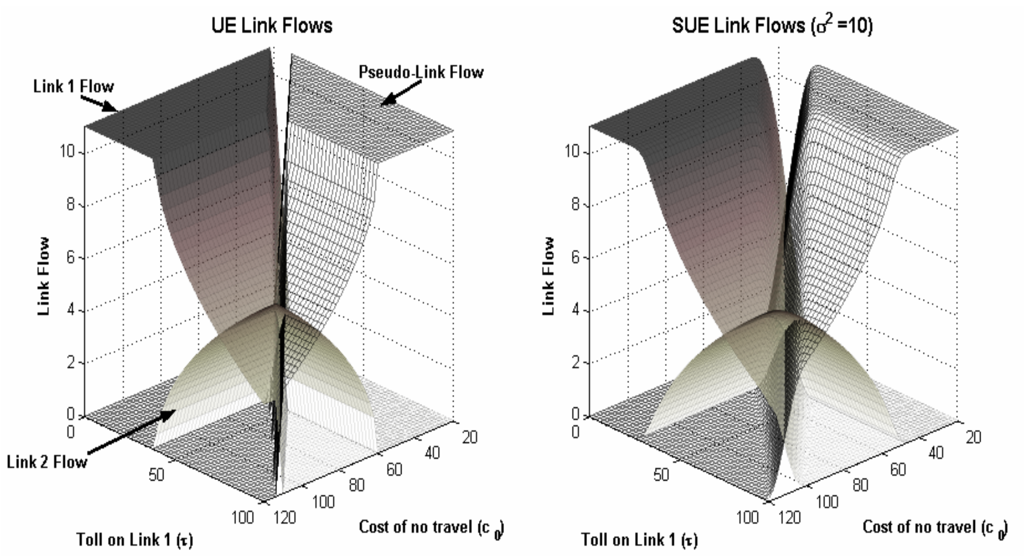


Figure 1: Variable Demand UE and SUE flows on the two-link network.

In addition to examining the link flows and their dependence on the underlying parameters of the network, other quantities can also be measured. For example, the revenue generated by the toll on link1 is simply $R = x_1\tau$. The dependence of revenue on the toll level and the OD demand for UE flows is shown in Figure 3. For the UE case (mesh surface) the flow on link1, and hence the revenue, drops to zero when the fixed cost on link 1 (free flow cost plus toll) increases beyond the cost of no travel (here $c_0 = 100$). Notice that the revenue surface with UE network flows has two local maxima, whereas with SUE flows there is a single global maximum for the revenue surface.

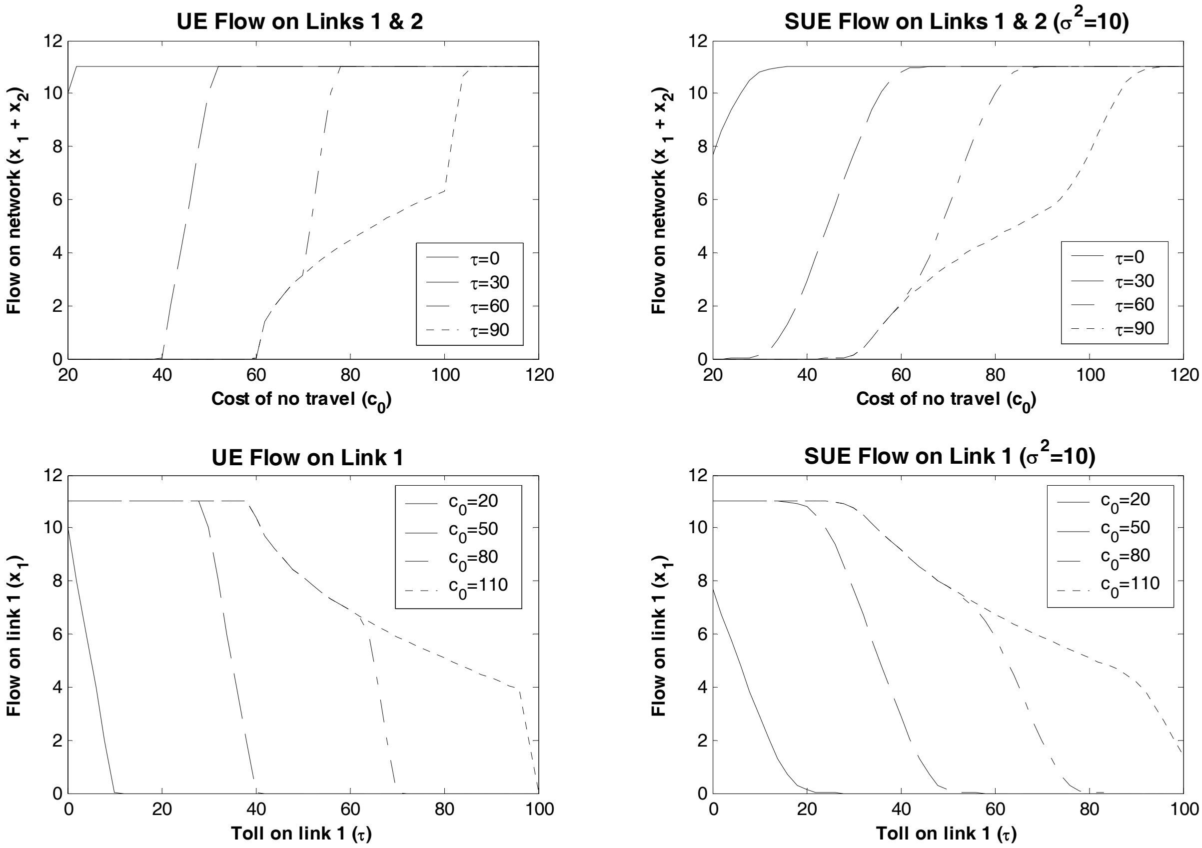


Figure 2: The UE and SUE flows on the two-link network as demand and toll are changed.

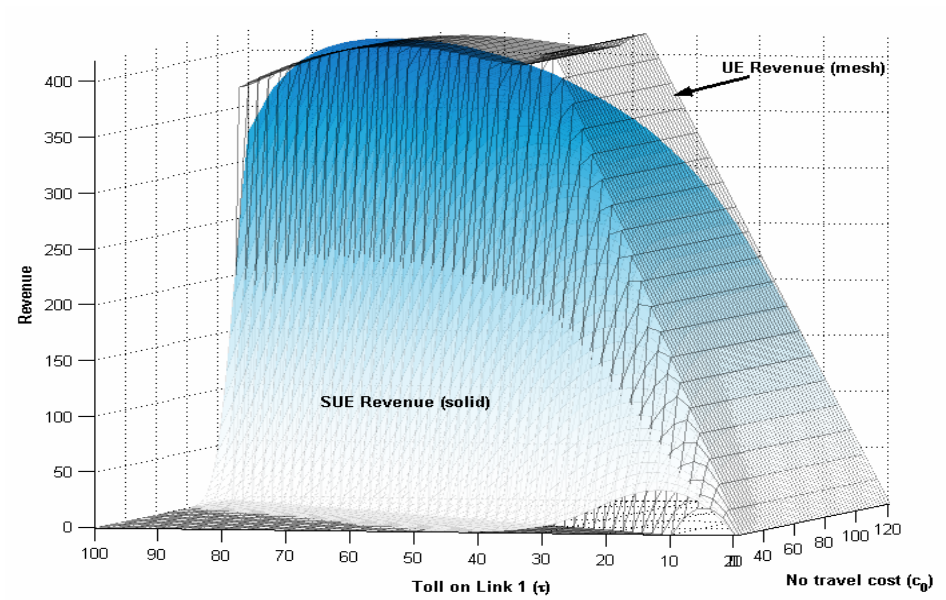


Figure 3: Revenue from Link 1 at UE (mesh) and SUE (solid) with $\sigma^2 = 10$

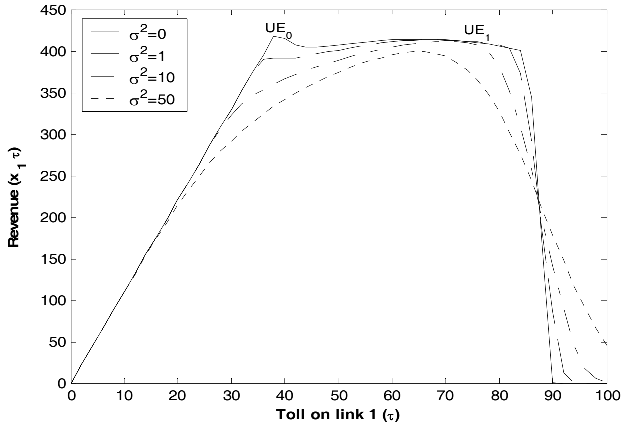


Figure 4: Revenue at fixed demand for UE and SUE with several variances

5.3 Implicit programming formulation of the NDP

The problem expressed in (6) can be simply reformulated as follows:

$$\max_{\mathbf{s}} \tilde{g}(\mathbf{s}) \quad \text{where } \tilde{g}(\mathbf{s}) \equiv g(\mathbf{s}, \mathbf{f}(\mathbf{s})) \tag{7}$$

Two properties of the probit SUE, including (i) the uniqueness path flows (and hence link flows) and (ii) the smoothness of the relationship between the path flows and design variables, allow the application of the implicit programming approach to (7). With the reformulated problem, various types of non-linear optimization algorithms can be applied directly to the problem provided that the Jacobian of the objective function is available.

A local linear approximation to the (surface of) SUE flows can be obtained via sensitivity analysis and this is sufficient to understand the changes in the SUE flows that result from small changes to the network design parameters (**s**). These approximate equilibrium flows can then be used to evaluate the objective function of the NDP, giving insight into the (local) consequences of changing the network design parameters.

For SUE with *fixed* demand, Clark & Watling (2002) derived an expression for the leading order linear changes in the link flows resulting from perturbations of the network design parameters (link cost function or OD demand changes).

The SUE path flows can be written as a function of the design parameters as

$$\mathbf{f}^* = \mathbf{f}^*(\mathbf{s})$$

with **f*** a solution to the variable demand SUE fixed point problem (5). The SUE flows depend on the network design parameters, **s**, that can include link cost parameters, OD demands and covariance matrix terms. The vector of path cost functions is then $\mathbf{c}(\mathbf{f},\mathbf{s})$. Consider the function

$$\mathbf{h}(\mathbf{f},\mathbf{s}) = \mathbf{f} - \mathbf{P}(\mathbf{c}(\mathbf{f},\mathbf{s}))\mathbf{q}\,.$$

For any given 'setting', **s**, of the design parameters, $\mathbf{h}(\mathbf{f}^*,\mathbf{s}) = \mathbf{0}$ where $\mathbf{f}^* = \mathbf{f}^*(\mathbf{s})$ are the corresponding SUE link flows.

The Clark & Watling (2002) result is that

$$\mathbf{f}^*(\mathbf{s}) \cong \mathbf{f}^*(\mathbf{s}_0) - \mathbf{J}_1^{-1}\mathbf{J}_2(\mathbf{s} - \mathbf{s}_0)\,, \tag{8}$$

where the Jacobian matrices ($\mathbf{J}_1, \mathbf{J}_2$) are the derivatives of $\mathbf{h}(\mathbf{f}(\mathbf{s}_0), \mathbf{s}_0)$ with respect to f and **s** respectively. Following section 4 this result for fixed demand SUE extends immediately to the case of variable demand SUE by extending the path set to include a constant cost pseudo-link for each OD pair.

With this sensitivity expression, we can calculate the derivative of path flows (**f**) respect to the design variables (**s**) as follows:

$$\nabla \mathbf{f} = \frac{\mathbf{f}^*(\mathbf{s}) - \mathbf{f}^*(\mathbf{s}_0)}{\mathbf{s} - \mathbf{s}_0} \cong -\mathbf{J}_1^{-1}\mathbf{J}_2 \qquad (9)$$

In calculating the Jacobian matrices ($\mathbf{J}_1, \mathbf{J}_2$), we also need the information on the Jacabian of path-choice probability. The method suggested by Daganzo (1979) is adopted in this paper to compute the jacobian of path-choice probability respect to the design variable.

One method of solving the NDP would be to apply the Sequential Quadratic Programming (SQP) algorithm to the problem in (7). At each outer iteration of the SQP, the algorithm demands information about the evaluation of the objective function of (7) and the Jacobian of the objective function (9). Thus, given a vector of design variable ($\mathbf{s}_k$) at iteration k, the optimization process re-calculates the SUE flows, evaluates the objective function at the new SUE flows, and calculates the Jacobian of the objective function respect to the design variable. The SQP algorithm then uses this information to determine the predicted optimal design variable ($\mathbf{s}_{k+1}$) for the next iteration.

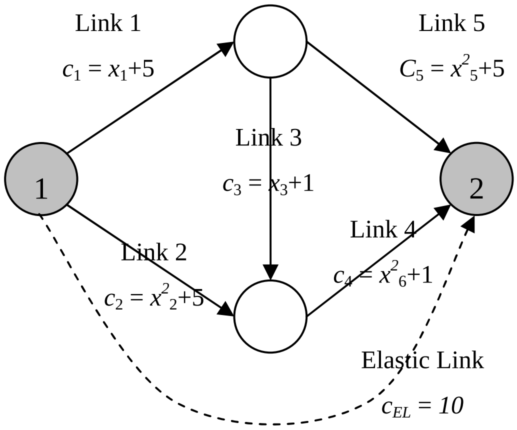
To ensure that the path flows and link flows change smoothly, one potential strategy for the MPEC would be to first solve a 'high-variance' SUE, before allowing the variances to return to their 'correct' level (zero for the UE solution). Figure 4 illustrates that this approach would lead to the solution UE_1 (finding the optimum of the high-variance SUE, then taking the limit $\Sigma \to 0$), whereas the global UE optimum occurs at the point UE_0. The processes of solving the NDP and taking the limit $\Sigma \to 0$ therefore seem to be non-commutative. Note in addition, that Figures 3 and 4 show the UE revenue to have multiple local optima whereas the SUE revenue has only one (global) optimum.

6. Numerical results with the five-link network

The algorithm developed in the previous section is tested with a simple five-link network. Figure 5 shows the five-link network with all link cost functions (including the elastic pseudo link). For the first test, we set the demand of 5 units wishing to travel from node 1 to node 2. The β is set to 0.3 for the initial test. The NDP considered in this test is the optimal toll problem with the objective of maximising the revenue.

Figure 6 illustrates the gradient of link flow on link 3 (x_3) respect to the toll on link 3 (τ_3). Two different gradient profiles are presented. The bold line is the gradient profile calculated by numerical differencing. The SUE link flows are calculated at each toll level and the numerical differencing gradient is calculated following:

$$\frac{x_3^*(\tau) - x_3^*(\tau_0)}{\tau - \tau_0}$$



$$\sigma = \beta \cdot \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0.5 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0.5 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

Figure 5: Five-link network

The dash line represents the gradient of link flow at each toll level calculated from the sensitivity expression, $-\mathbf{J}_1^{-1}\mathbf{J}_2$. The gradient of link flow as calculated by the sensitivity analysis expression is relatively smoother than the gradient calculated from numerical differencing. Based on the information on the link flow gradient, the approximated link flow can be calculated as the function of the link toll, $x_3(\tau) = x_3(\tau_0) - \mathbf{J}_1^{-1}\mathbf{J}_2 \cdot (\tau - \tau_0)$. Figure 7 shows the linear approximation of the flow on link 3 as the function of the link toll at three different initial toll levels ($\tau = 0,4,10$). The bold line is the real SUE link flow. As expected, the linear approximations of the link flow in all cases closely estimate the real link flow around the neighbourhood of the initial toll level. The approximated link flows are all linear relationships respect to the link toll since the linear approximation is adopted.

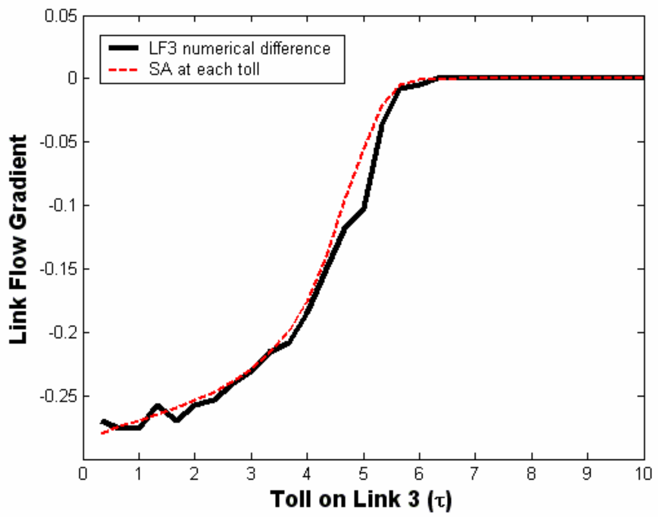


Figure 6: Link flow gradient from sensitivity analysis and numerical differencing

The direct product of the ability to approximate the link flow from the initial SUE link flow is the derivative of the objective function of the NDP. In this test, a simple objective function of maximising the revenue is:

$$g = \sum_j x_j \cdot \tau_j$$

and assuming that the toll is only imposed on link 3, the derivative of the objective function respect to the link toll can be defined as:

$$\frac{\partial g}{\partial \tau_3} = \tau_3 \cdot \frac{\partial x_3}{\partial \tau_3} + x_3 .$$

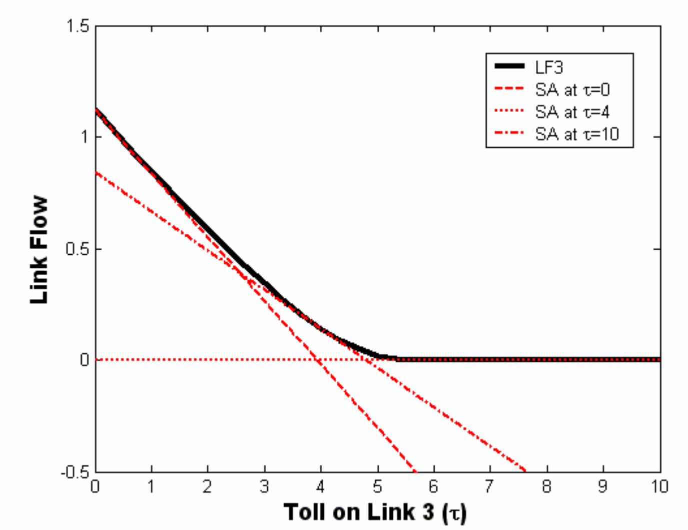


Figure 7: Linear approximation of link flow respect to the toll on link 3

Figure 8 shows the gradient of the revenue curve respect to the toll on link 3 calculated by sensitivity analysis expression and numerical differencing. Similar to the case of the link flow gradient, the gradient of the revenue curve as calculated by the sensitivity analysis expression is much smoother than the gradient from the numerical differencing method.

As illustrated above, the derivative of the link flow respect to link toll is a linear function of the toll level. Thus, the approximated revenue at each toll level will be a quadratic function of the toll level. Figure 9 depicts the real revenue curve (bold line) and the approximated revenue curve at different initial toll levels of 0, 4, and 10. At each iteration of the SQP, the information on the gradient of the revenue curve and approximated revenue curve will be used to define the predicted optimal toll level for the next iteration. For example, from Figure 9 if the initial iteration starts at the toll level of zero, the SQP will seek the optimal toll level for the approximated revenue curve which will be about 2. The toll level of 2 will then become the toll level for the next iteration.

With this operation, the NDP with Probit SUE can be solved directly with the SQP. The first example presented in this section is to optimize a single toll level on link 3 in order to maximize the revenue. Figure 9 shows the optimization process from the SQP algorithm. The bold line represents the real revenue curve, the dots represent the iteration of SQP, and the dash lines represent the approximated revenue curves at each of the SQP iterations. The initial toll level is set to be at 0.2. As mentioned, the SQP received the information about the gradient of link flow and revenue curve respect to the toll on link 3, then the predicted optimal toll is calculated from the approximated revenue curve (dash curve). The predicted toll is 2.04. Then, the new SUE flow is calculated for the toll level of 2.04. The information on the gradient of the link flow and revenue curve at the toll level of 2.04 is, again, passed to the SQP algorithm which predicts the next toll level of 2.25. The algorithm terminates here meaning that the toll level of 2.25 is the optimal toll level for link 3.

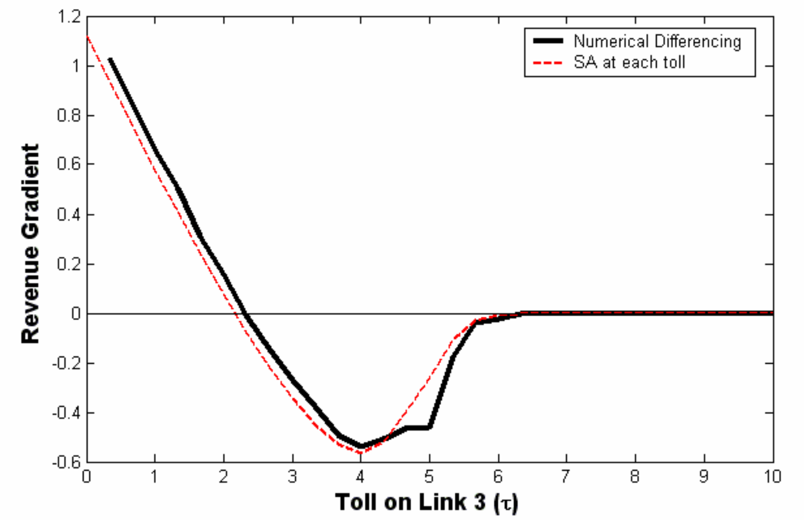


Figure 8: Revenue curve gradient from sensitivity analysis and numerical differencing

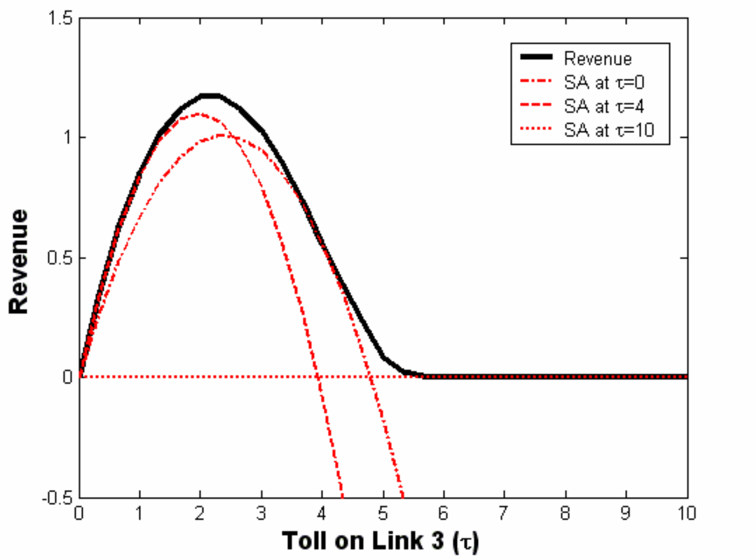


Figure 9: Approximation of revenue respect to the toll on link 3

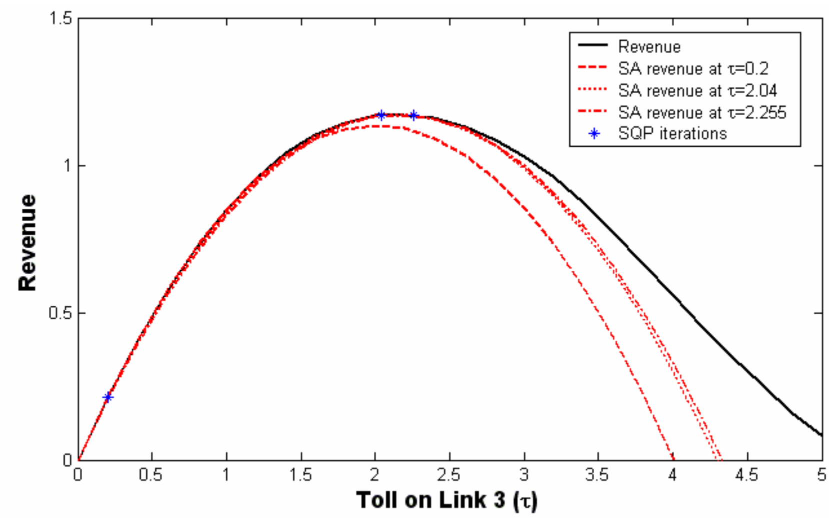


Figure 10: SQP iterations for optimizing the toll on link 3

The next test is to find the optimal tolls for link 3 and link 5. Similarly to the case of single tolled link, an approximated revenue surface is created based on the initial flow condition, then the SQP predicts the optimal tolls based on this approximated surface. Figure 11-13 show the approximated revenue surface from three different initial points. The mesh-grids in Figure 11-13 are the real revenue surface where the dark surfaces are the approximated revenue surfaces.

From the real revenue surface, we can observe the existence of multiple-optima in three different regions. The first region is the region with the 'dome shape' revenue surface. In this region, both link 3 and link 5 still have some link flows and the global optimum is in this region. Figure 11 shows the surface approximating the first region that is approximated at the toll levels of 0.4 and 0.5 on link 3 and link 5 respectively. The second region is the curve on the left hand side of the dome shape. In this region, there is no traffic volume on link 5 since the toll on this link is too high (started from about 10). If the initial point for the approximation is in the vicinity of this second region, the approximated revenue surface will look like the curve shown in Figure 12 (the initial point for the approximation are the toll levels of 0.4 and 11 on link 3 and 5 respectively). The final region is on the right hand side of the dome shape where there is no link flow on link 3 (toll on link 3 starts from about 10). The toll on link 3 does not effect the revenue of in this region. Figure 13 shows the approximated revenue surface of this last region which is approximated at the toll levels of 9 and 5 on link 3 and 5 in that order.

Figure 14 shows the contour of the revenue surface as the function of the toll levels on link 3 and 5. The figure also shows the optimization process (from the SQP) started from three different initial solutions. Table 1 shows the initial solution and the final solution converged for each optimization test.

Table 1: Initial points and solutions for optimization A, B, C for the tolls on link 3 and 5

Test	Initial solution (toll on link 3, toll on link 5)	Solution (toll on link 3, toll on link 5)	Revenue
A	0.40, 11.00	3.53, 10.96	2.45
B	9.00, 5.00	3.28,3.28	4.00
C	0.40,0.50	3.64, 3.41	3.99

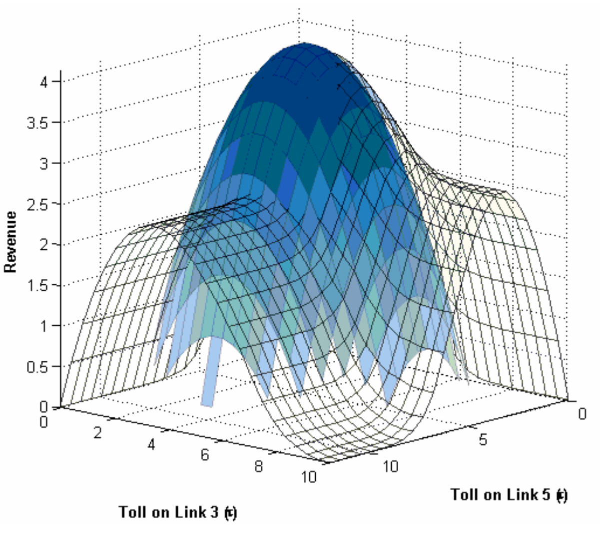


Figure 11: Approximated revenue surface at the toll level of 0.4 and 0.5 on link 3 and 5

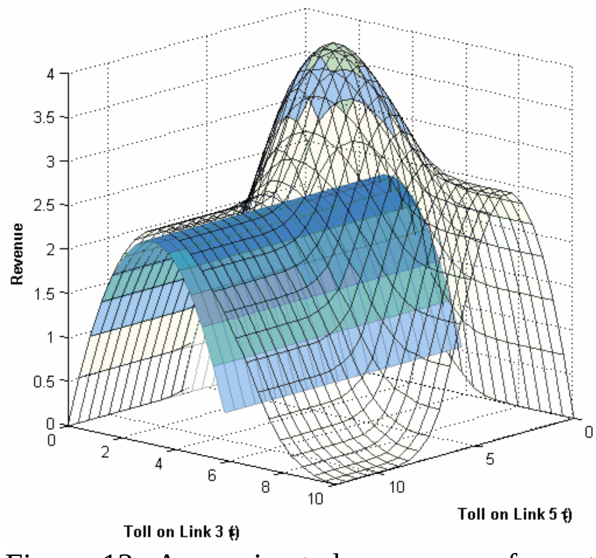


Figure 12: Approximated revenue surface at the toll level of 0.4 and 11 on link 3 and 5

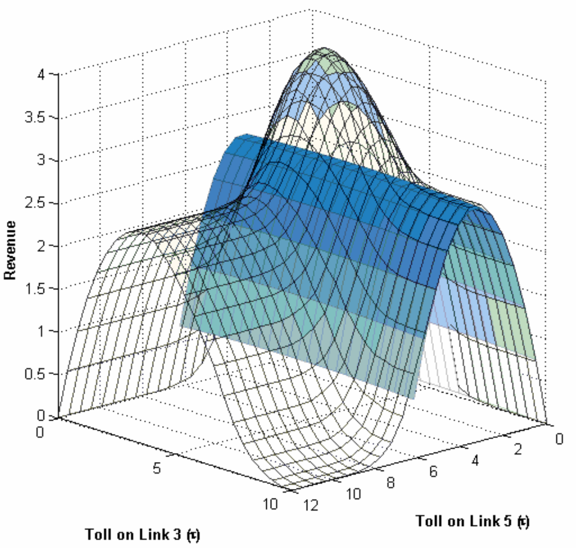


Figure 13: Approximated revenue surface at the toll level of 9 and 5 on link 3 and 5

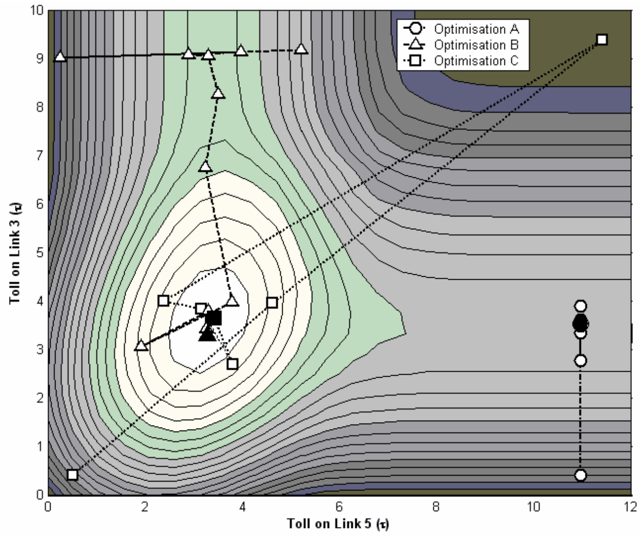


Figure 14: Optimization process for the toll level on link 3 and 5 starting from three different points

The optimization A started at the region that has the approximated surface shown in Figure 12. The optimization process is trapped at the local optimum in this region. The final solution is the local optimal with the toll levels of 3.53 and 10.96 for link 3 and link 5 respectively. The revenue from this solution is about 2.45. In the optimization B, the SQP is started at the toll levels of 9 and 5 on link 3 and 5. Figure 13 shows the approximated revenue curve from this starting point. Despite the similarity between the starting region of the optimization A and B, the optimization process found its ways to the global optimum in the dome-shape region. This is because there exists some slope toward the dome-shape region on the top of the approximated revenue curve shown in Figure 13. This is not the case for the approximated curve in Figure 12. Finally, the optimization C is started at the point near the no-toll case (0.4 and 0.5). The approximated revenue curve at this point is shown in Figure 11. This starting point in indeed in the region with the global optimum. As expected, the final solution of the optimization C converged to the global optimum at the toll levels of 3.64 and 3.41 on link3 and 5 respectively.

Overall, the optimization process seems to work well in solving the NDP with the Probit SUE, at least for the simple case. However, there are various implementation issues that may pose some problem on the optimization process. The first one is the lack of the smoothness of the objective function surface. This is mainly due to the lack of the convergence of the MSA (Method of Successive Average) adopted to find the Probit SUE. Figure 15 shows the detail surfaces of the revenue curves (toll on link 3) with different numbers of MSA iterations.

Three different numbers of MSA iterations were tested, including 10000, 1000, and 500 iterations. As expected, the revenue curve with the lowest number of MSA iterations (500 iterations) is relatively non-smooth compared to the revenue curve with a higher number of MSA iterations. The reason is due to the convergence error of the SUE flows. This is also the reason for the non-smoothness of the gradient of the link flow calculated by the numerical differencing method (see Figure 6 above). It is important to address this issue before attempting to apply this method to a large scale network problem(which may involve a higher magnitude of the convergence error).

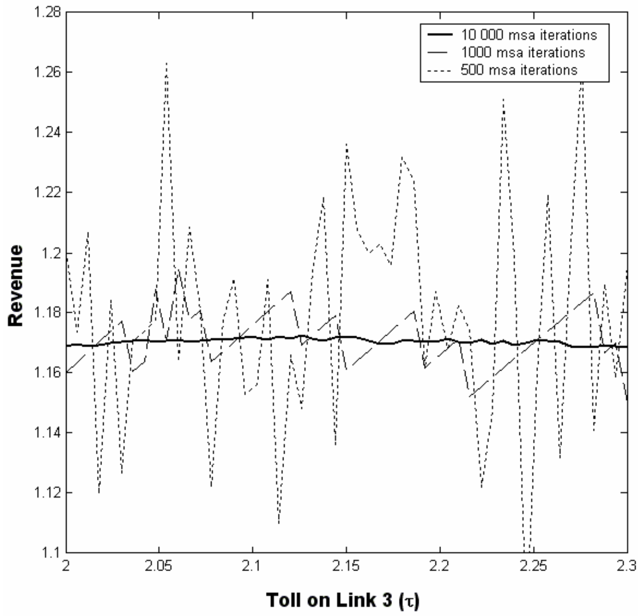


Figure 15: Comparison of the revenues curves generated by different MSA iterations

The other source of non-smoothness is, as mentioned earlier, due to the level of the variance of the Probit model. As described earlier, the nature of UE (which can be seen as the SUE case with zero variances) is likely to exhibit some non-smooth changes of the link flows (due to the change of active paths set). Figure 16 shows the comparison of the revenue curve against the different constant values of the pseudo link cost (with toll level on link 3 =1) with three different levels of β, including 1, 0.3, and 0 (UE case).

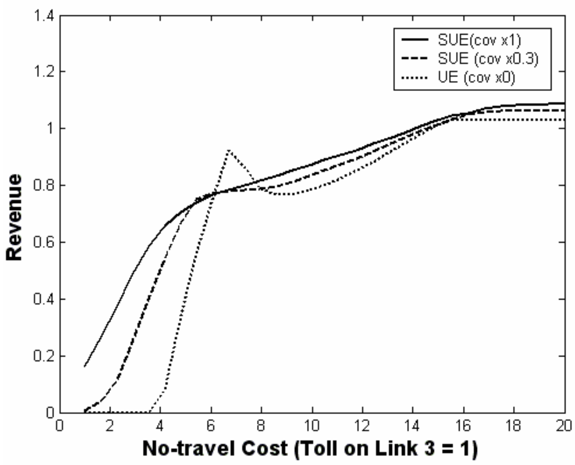


Figure 16: Comparison between the revenue curves with different values of variance

The revenue curves with $\beta = 1$ and 0.3 (SUE case) are reasonably smooth. On the other hand, with the UE case ($\beta = 0$) the revenue curve exhibits the 'cusp' point (around the constant value of 6) which may be non-differentiable. This point happened when there was a change of the set of used paths. This result confirms the advantage of utilising SUE as the equilibrium rule for the NDP. Nevertheless, there still some possibility that the NDP with the Probit SUE may involve some non-differentiable points. This can be happen in the case with very low variances. This is also possible from the practical point of view, since in theory all paths are used by the SUE but in practice this is not the case. Due to some computational precision, some paths with very low path choice probabilities may not be active paths. This condition may be changed when the design variable is changed and the unused paths become active (exactly the same case with the UE). Further investigation on this point is necessary to ensure the functionality of the method proposed with a larger scale problem.

7. Conclusions & further research

In this paper we have considered the Network Design Problem (NDP) in the case where the user and network response is governed by an elastic demand, Probit SUE model. Such a model has some claims to greater behavioural realism than the commonly used UE model, yet here we have particularly focused on the properties it induces in the NDP. In particular it has been seen how the perceptual variance has desirable consequences in terms of the smoothness of the resulting NDP, implying that gradient methods such as sensitivity analysis may be used to infer sensible local search directions. In this paper, the aim has been to understand the nature of the NDP, rather than solve it, but given the results from the simple example illustrated, it is a natural next step to the explore the solution of such problems via the use of sensitivity analysis, building on the promise of such techniques from previous 'bilevel' applications of this nature (e.g. Yang, 1995, 1997; Patriksson & Rockafellar, 2003).

The examples studied in this paper were deliberately chosen to be simple, to serve illustration of the principles, yet it would clearly be valuable to conduct further investigations of the comparisons conducted on larger scale networks. This would serve to provide greater evidence of the relationship between the NDP-UE and NDP-SUE responses, and allow the further investigation as to what extent NDP-SUE may be viewed as a smoothing method approximation for solving NDP-UE problems, and to what extent the perceptual variance term impacts on the existence and multiplicity of local optima. Given the generality of the SUE fixed point formulation, we may also examine asymmetric SUE problems when the 'lower level' may no longer be a singleton (multiple separated SUE). The other key practical issues that must be considered for developing a method for a large scale network are the issue of convergence error from the MSA and its impact on the calculation of the link flow gradient.

Acknowledgements

The research described in this paper was conducted under the financial support of the U.K. Engineering & Physical Sciences Research Council.

Thanks to John Nellthorp for contributions to our discussions on this work.

References

Beckmann, M.J., McGuire, C.B. and Winsten, C.B., 1956. Studies in the Economics of Transportation. Yale University Press, New Haven, Conn.

Ben-Akiva, M., De Palma, A. and Kanaroglu, P., 1986. Dynamic Model of Peak Period Traffic Congestion with Elastic Arrival Rates. Transportation Science 20(2) 164-181.

Brotcorne, L., Labbé, M., Marcotte, P., and Savard, G., 2001. A Bilevel Model for Toll Optimization on a Multicommodity Transportation Network. Transportation Science 35(4) 345-348.

Cantarella, G.E. and Binetti, M.G., 2002. Stochastic Assignment with Gammit Path Choice Models. Transportation Planning: State of the Art, eds. M. Patriksson & M Labbé, Kluwer, Dordrecht, Netherlands, 53–68.

Cantarella, G.E. and Cascetta, E., 1995. Dynamic Processes and Equilibrium in Transportation Networks: Towards a Unifying Theory. Transportation Science 29(4) 305-329.

Clark, S.D. and Watling, D.P., 2002. Sensitivity analysis of the probit-based stochastic user equilibrium assignment model. Transportation Research 36B, 617-635.

Daganzo, C., 1979. Multinomial Probit: The Theory and Its Application to Demand Forecasting, Academic Press Inc, New York.

Davis, G.A., 1994. Exact local solution of the continuous network design problem via stochastic user equilibrium. Transportation Research 28B, 61-75.

Fisk, C., 1980. Some Developments in Equilibrium Traffic assignment. Transportation Research 14B(3) 243-255.

Gentile, G. and Papola, N., 2001. Network design through sensitivity analysis and singular value decomposition. Paper presented at TRISTAN IV, San Miguel, Azores, June 13^{th}–19^{th} 2001

Larsson, T., Lundgren, J.T., Patriksson, M. and Rydergren, C., 2001. Most likely traffic equilibrium route flows—analysis and computation. Equilibrium Problems & Variational Methods: International Workshop in Memory of Marino De Luca, Taormina, Italy, December 1998.

Luo, Z.Q., Pang, J.S. and Ralph, D., 1996. Mathematical Programs with Equilibrium Constraints. Cambridge University Press.

Maher, M.J., Hughes, P.C. and Kim, K.S., 1999. New algorithms for the solution of the stochastic user equilibrium assignment problem with elastic demand. Proceedings of the 14^{th} International Symposium on Transportation and Traffic Theory, Jerusalem, Israel.

Nielsen, O.A., Daly, A. and Frederiksen, R.D., 2002. A Stochastic Route Choice Model for Car Travellers in the Copenhagen Region. Networks and Spatial Economics 2(4) 327-346.

Patriksson, M. and Rockafellar, R.T., 2002. A Mathematical Model and Descent Algorithm for Bilevel Traffic Management. Transportation Science 36(3) 271-291.

Patriksson, M. and Rockafellar, R.T., 2003. Sensitivity Analysis of Aggregated Variational Inequality Problems, with Application to Traffic Equilibria. Transportation Science 37(1) 56–68.

Prashker, J.N. and Bekhor, S., 1999. Stochastic User-Equilibrium Formulations for Extended-Logit Assignment Models. Transportation Research Record 1676, 145-151.

Sheffi, Y., 1985. Urban Transportation Networks. Prentice Hall, New Jersey.

Sheffi, Y. and Powell, W.B., 1981. A Comparison of Stochastic and Deterministic Traffic Assignment over Congested Networks. Transportation Research 15B(1) 53-64.

Smith, M.J., 1979. The Existence, Uniqueness and Stability of Traffic Equilibria. Transportation Research 13B, 295-304.

Tobin, R.L. and Friesz, T.L., 1988. Sensitivity Analysis for Equilibrium Network Flow. Transportation Science 22(4) 242-250.

Yang, H., 1995. Heuristic algorithms for the bilevel origin-destination matrix estimation problem. Transportation Research 29(4) 231-242.

Yang, H., 1997. Sensitivity analysis for the elastic-demand network equilibrium problem with applications. Transportation Research 31B(1) 55-70.

Yang, H. and Bell, M.G.H., 1998. Models and algorithms for road network design: a review and some new developments. Transport Reviews 18, 257-278.

Yang, H. and Bell, M.G.H., 2001. Transport bilevel programming problems: recent methodological advances. Transportation Research 35B(1) 1–4.