

19th International Symposium on Transportation and Traffic Theory

A splitting rate model of traffic re-routing and traffic control

Mike Smith^{a,*}, Richard Mounce^b

^a*Department of Mathematics, University of York, Heslington, York, YO10 5DD, UK*

^b*Department of Architecture, University of Cambridge, 1-5 Scroope Terrace, Cambridge, CB2 1PX, UK*

Abstract

The paper presents an idealised dynamical model of day-to-day or within-day re-routing using splitting rates at nodes, or node-exit flows, rather than route-flows. It is shown that under certain conditions repeated application of the dynamical model gives rise to a sequence of link flow vectors which converges to a set of approximate Wardrop equilibria. A special dynamical signal green-time re-allocation model is added; the combination is shown (in outline) to converge to the set of approximate consistent equilibria under certain conditions. Finally the paper uses model network results to illustrate a method of designing fixed time signal timings to meet different scenarios. The paper sets the scene by giving a simple motivating model, involving both routing changes and green-time changes following the equisaturation policy, in which unpredictable discontinuous behaviour, including a pitchfork bifurcation, occurs as a result of the interaction between re-routing and responsive control.

© 2011 Published by Elsevier Ltd. Open access under [CC BY-NC-ND license](https://creativecommons.org/licenses/by-nc-nd/4.0/).

Keywords: Type your keywords here, separated by semicolons ;

1. Introduction

This paper considers idealised models of both routing changes as drivers seek better routes (either from day to day or within a day) and signal control changes as an adaptive control system reacts to changing link flows and hence routing decisions (either from day to day or within a day). The main emphasis here is to combine within a single theoretical framework the modelling of routing and signal control changes, seeking (1) to understand routing – control interactions and (2) to recognise or design beneficial new responsive and fixed time control systems for use in real networks. The paper uses several simple networks to illustrate the ideas. In one of these the design of intelligent fixed time control settings for networks subject to changed conditions, while taking reasonable account of route choices, is illustrated.

1.1 Re-routing models

Wardrop (1959) is credited with the idea of traffic equilibrium although the central idea goes back a long way; at a Wardrop equilibrium, for each origin-destination (OD) pair there is no flow on any more costly route. An

* Corresponding author. Tel.: +44 1904 32 4166; fax: +44 1904 32 3071.

E-mail address: mjs7@york.ac.uk.

equilibrium notion such as this is related to many possible dynamical systems and it is natural to seek such dynamical systems which have some realism (as day to day dynamical systems for example). Furthermore, if for example a day-to-day or within-day dynamical model converges to equilibrium quickly, then that dynamical model may also be used to estimate the equilibrium state or the long run network performance. This can be achieved simply by running the day to day system for many “days”. *The same model may perform two functions.*

It would be interesting if a reasonable, if idealized, day to day adjustment or reassignment process also converged *rapidly* to equilibrium, as in that case the possibility would arise of utilising the same model in two distinct ways. Both

- as an idealized day to day or even possibly a within day model suitable for helping to evaluate ITS interventions in the short term and also
- as a long term model suitable for helping the long term evaluation of ITS and other interventions and also alternative long term planning options.

There would then be software and/or algorithmic elements common to the short term model and the long term model and importantly, there would also be a natural consistency between certain long run and short run model outputs; as these outputs will have been produced by the same underlying model. It might seem that this is too much to wish for; yet there is now tantalising evidence supporting the above possibility. This possibility is supported by results generated by Gentile (2009) and Gentile and Noekel (2009), which suggest that a splitting rate assignment algorithm implemented within VISUM software can (at least in some cases) converge toward Wardrop equilibrium very quickly indeed. This paper was motivated by a consideration of the papers by Gentile (2009), Gentile and Noekel (2009) and Bar Gera (2002) and suggests a particular dynamical model, based on splitting rates at nodes, that may represent, albeit in an idealised manner, day to day and even within-day trip reassignment and may also be used to estimate (efficiently) long run Wardrop equilibria.

There have been very many proposed idealised models of day to day re-routing and the one which we begin with in this paper was proposed in Smith (1984a). In this model, for each pair of routes joining the same OD pair, traffic flow swaps from the more costly route to the less costly route at a rate which is proportional to the flow on the more costly route multiplied by the difference in costs between the two routes. Smith (1984a) shows that this simple and apparently natural dynamical route-swap system converges to equilibrium under rather weak conditions. *However He et al (2010) highlight a behavioural deficiency of this simple dynamical model.* This deficiency is described below.

In this paper we propose a *splitting rate* version of the above dynamical route-swap model, in which exit-link flows swap at each node following a rule similar to that above. At each node flow-weighted costs to each destination via the alternative exit links drive the dynamical system and exit flows swap only between different exit links at the same node. This splitting rate re-routing model does not exhibit the anomaly identified by He et al (2010). Furthermore it is shown in this paper that, under certain conditions (which preclude looping paths), the resulting dynamical model converges to approximate equilibrium.

1.2 Re-routing and control

The dynamical splitting rate traffic re-routing model is enhanced by adding a dynamical green-time (or red time) re-allocation model, following the responsive P_0 signal control policy. It is shown in outline that the combined dynamical system (now embracing varying green-times or red-times as well as varying traffic flows) also converges to equilibrium under certain conditions. This is of course a desirable property of such a joint dynamical system; it would not be good if an adaptive control system, when interacting with reasonable routing changes, creates an unstable or unpredictable system.

A context is provided by giving an initial example in which the equisaturation policy gives rise to a pitchfork bifurcation in a simple network. In this example the behaviour of the network *will* be unstable and unpredictable; entirely as a result of the combination of a particular reasonable signal control policy and reasonable route choice changes. The simple example also shows that there can be no provable stability results involving the equisaturation policy similar to the “global” stability result outlined here when the P_0 control policy is used.

Finally, the paper shows how the dynamic flow / control model here may be used to design signal settings suitable for particular situations; these signal settings take some account of traveller re-routing.

2. Network Control Using Signals: Equisaturation, Instability and a Pitchfork

In the signal controlled network shown in figure 1 below, both routes have

- the same free flow travel time and
- the same saturation flow of s vehicles per minute at the signal.

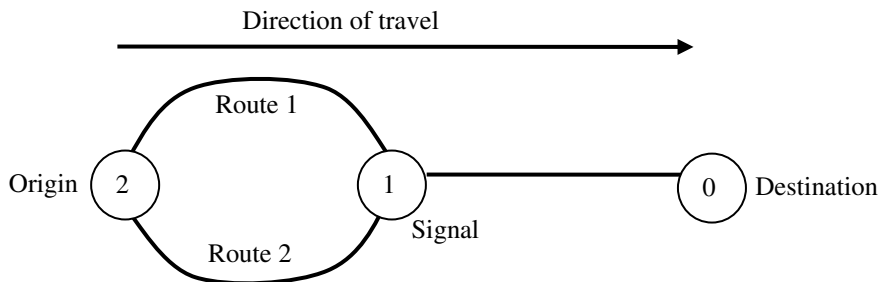


Figure 1. The signal at node 1 is rapidly adjusted to equalise the two degrees of saturation for any flows along the two routes. The network is symmetrical.

Let T be the total flow rate (vehicles per minute) from the Origin to the Destination via the two routes and let $0 < T < s$. Let H_1 be the proportion of this Origin-Destination flow which travels along route 1 and H_2 be the proportion travelling along route 2; so that $H_1 + H_2 = 1$. The flows on routes 1 and 2 will then be TH_1 and TH_2 . We will consider T fixed but also we will consider different values of T (satisfying $0 < T < s$). The delay formulae used impose strict capacity constraints.

Let G_1 be the green time proportion awarded to route 1 and G_2 be the green time proportion awarded to route 2, at the signal at node 1. Throughout this section we suppose that the signal quickly equalises the two saturation ratios at the signal; so $H_1, H_2, G_1,$ and G_2 must satisfy $H_1T/sG_1 = H_2T/sG_2$ or $H_1/G_1 = H_2/G_2$. Since G_1 and G_2 are green time proportions and so add to 1 it now follows that G_1 and G_2 must be given (for any T) by:

$$G_1 = H_1 \text{ and } G_2 = H_2. \tag{1}$$

(We assume that the timings are adjusted quickly.)

Throughout this section travel cost = travel time (in minutes per vehicle). We further assume that the delay costs d_1 and d_2 felt at the signal by vehicles traversing routes 1 and 2 are determined by putting:

$$d_1 = \frac{BTH_1}{sG_1(sG_1 - TH_1)} \tag{2}$$

and

$$d_2 = \frac{BTH_2}{sG_2(sG_2 - TH_2)} \tag{3}$$

and that the travel costs C_1 and C_2 along routes 1 and 2 (in minutes per vehicle) are then given as follows:

$$C_1 = K + ATH_1 + d_1 \tag{4}$$

and

$$C_2 = K + ATH_2 + d_2. \tag{5}$$

Here K, A and B are constants. Delay costs (2) and (3) experienced at the signal (minutes per vehicle) are identical to the second term of Webster’s delay formula if B is chosen to be $9/20$. (Webster’s formula is designed for Poisson traffic.) So d_1 and d_2 will be the estimated delays at the junction and $K + ATH_1$ and $K + ATH_2$ are the running times (minutes / vehicle) along the remainder of the two routes. Whole journey costs are obtained by adding these two components as in (4) and (5) above. It follows immediately, using (1), that

$$d_1 = \frac{BTH_1}{sH_1(sH_1 - TH_1)} \text{ and } d_2 = \frac{BTH_2}{sH_2(sH_2 - TH_2)}. \tag{6}$$

To find equilibria (assuming that the equisaturation policy is followed) where route costs are equal and so no traveller has any incentive to change route it is natural to solve the equation $C_1 - C_2 = 0$. This equation may be written (using (4), (5) and (6)):

$$C_1 - C_2 = ATH_1 - ATH_2 + \frac{BTH_1}{sH_1(sH_1 - TH_1)} - \frac{BTH_2}{sH_2(sH_2 - TH_2)} \tag{7}$$

$$= AT(H_1 - H_2) - \frac{BT}{s(s-T)} \frac{H_1 - H_2}{H_1H_2} = 0. \tag{8}$$

It is clear from (7) and (8) that, for any T , $H_1 = H_2 = \frac{1}{2}$ yields one solution of $C_1 - C_2 = 0$ and so yields an equilibrium (consistent with the equisaturation policy): but are there others?

Suppose we have another solution \mathbf{H} to (8). This solution must have $H_1 - H_2 \neq 0$; so, dividing (8) by $T(H_1 - H_2)$,

$$A - \frac{B}{s(s-T)H_1H_2} = 0 \tag{9}$$

and hence:

$$H_1H_2 = \frac{B}{As(s-T)}. \tag{10}$$

Putting $H_2 = 1 - H_1$ in (9) or (10) and solving the resulting quadratic in H_1 we obtain two *possible* new equilibrium solutions; these are:

$$H_1 = \frac{1}{2} - \left(\frac{1}{4} - \frac{B}{As(s-T)} \right)^{\frac{1}{2}} ; H_2 = \frac{1}{2} + \left(\frac{1}{4} - \frac{B}{As(s-T)} \right)^{\frac{1}{2}} \tag{11a}$$

and

$$H_2 = \frac{1}{2} - \left(\frac{1}{4} - \frac{B}{As(s-T)} \right)^{\frac{1}{2}} ; H_1 = \frac{1}{2} + \left(\frac{1}{4} - \frac{B}{As(s-T)} \right)^{\frac{1}{2}}. \tag{11b}$$

These really are two additional solutions of $C_1 - C_2 = 0$ if all four expressions in (11) are real and also differ from the already known solution $H_1 = \frac{1}{2}$ and $H_2 = \frac{1}{2}$. This happens if and only if

$$\frac{1}{4} - \frac{B}{As(s-T)} > 0 \tag{12}$$

so that the square roots in (11) are real and non-zero. Condition (12) may be equivalently written:

$$T < s - \frac{4B}{As}. \tag{13}$$

The right hand side of inequality (13) occurs as a label on the TH_i axis in figure 2. The set of all equilibria (TH_1 , TH_2) as T varies from 0 to s is shown by bold lines in figure 2. [In fact each ‘‘axis’’ of the figure may also be regarded as comprising equilibria, and so these axes are bold too. The analysis here and the picture make the assumption that $s-4B/As > 0$.]

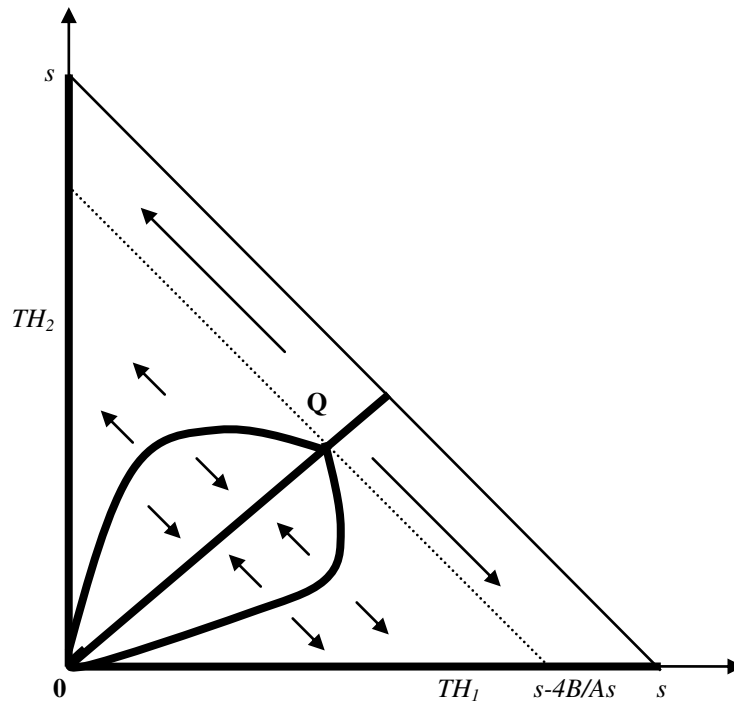


Figure 2. Bold lines indicate the set of route-flow vectors $\mathbf{TH} = (TH_1, TH_2)$ which are equilibria consistent with the equisaturation policy for some T .

In figure 2, five branches of the equilibrium set, three of them comprising stable equilibria, converge on the origin as T becomes small. Arrows indicate the directions of motion of non-equilibria, for various values of T , as drivers and green-time respond to delays and saturation ratios: drivers seeking cheaper routes and green-times seeking equisaturation. (See 2.1 below.)

Assuming that the directions of these arrows are correct, let us suppose that T slowly increases from zero. If the flow pattern initially follows the central (stable) equilibrium set leading away from the origin then it will carry on doing that until $T = s - 4B/As$, when the bifurcation point Q is reached. As T increases beyond this point the central equilibria become unstable and the flow pattern will then be likely to wobble and (falling off the now unstable set of central equilibria), follow one of the long arrows (with flow swapping from more to less costly routes) and continually move toward one of the two axes. Thus the behaviour of the system changes sharply near Q . This is in part due to the change in the topological structure of the set of equilibria at this point. Consider a small circle around Q : for $T < s - 4B/As$ the equilibrium set has three points within the circle and for $T > s - 4B/As$ the equilibrium set has just one point within the circle. The topological structure of the equilibrium set changes at Q : this point is called a bifurcation. The shape of the equilibrium set near Q is that of a pitchfork.

2.1 Verification of the direction of motion at non-equilibria in this example

Suppose T is fixed and \mathbf{H} is a non-equilibrium route proportion vector; so that costs on the two routes are unequal. Then it is natural to suppose that traffic swaps from the more to the less costly route. It is necessary to check that the directions of the arrows in figure 2 accord with this. By (7) and (8),

$$C_1 - C_2 = AT(H_1 - H_2) - \frac{BT}{s(s-T)} \frac{H_1 - H_2}{H_1 H_2} = \left(A - \frac{B}{s(s-T)} \frac{1}{H_1 H_2} \right) T (H_1 - H_2).$$

Consider just the interior of the lower triangle in figure 2 adjacent to the horizontal or TH_1 axis; in this region

$$T(H_1 - H_2) > 0$$

and so, dividing by $T(H_1 - H_2)$, $C_1 - C_2$ has the same sign as

$$A - \frac{B}{s(s-T)} \frac{1}{H_1 H_2}. \tag{14}$$

We are assuming that if this is positive then, since $C_1 > C_2$, flow will swap from route 1 to route 2 and if this is negative then, since $C_1 < C_2$, flow will swap from route 2 to route 1. This expression (14) is certainly negative if $H_1 H_2$ is sufficiently small since $\frac{B}{As(s-T)}$ is, for any T with $0 < T < s$, a positive constant independent of \mathbf{H} . Now at any point near the horizontal axis H_2 is small and therefore $H_1 H_2$ is indeed small. In this case $C_1 - C_2 < 0$ and traffic will swap from route 2 to route 1 as indicated by those arrows pointing toward the horizontal axis in this lower triangular region. A similar analysis applies elsewhere where $C_1 - C_2 \neq 0$, and justifies all the arrows in figure 2 indicating the direction of motion of non-equilibrium flows.

3. An apparently reasonable model of the route adjustment process

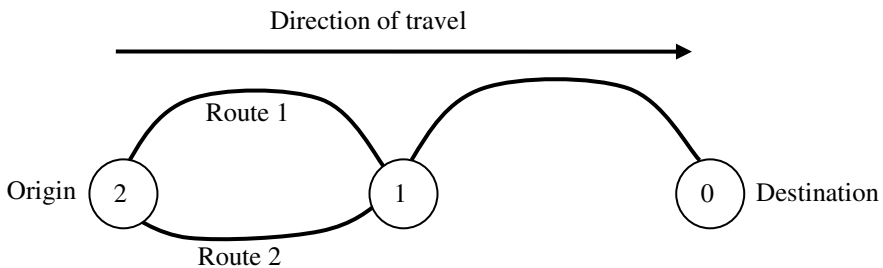


Figure 3. A two route network

Smith (1984a) proposed a simple and reasonable-looking day to day re-routeing process; the purpose was three-fold:

1. to allow the stability or otherwise of a given traffic equilibrium to be studied as unstable equilibria are unlikely to persist;
2. to determine, when there are several equilibria, which equilibrium attracts trajectories starting from a given point; and
3. to allow the possibility of modelling moving a traffic equilibrium to another “better” equilibrium by using a perhaps temporary signal-control intervention.

Suppose that travellers traverse the small network illustrated in figure 3, and that they do this repeatedly, day after day. Suppose that both routes are used and that currently (on day t) route 1 is more costly than route 2. How many travellers will swap from route 1 to route 2 on day $t + 1$? What are the alternatives? Here we give some suggestions and extend those suggestions to deal with more complicated networks.

In this simple case, at first sight the simplest assumption is that some travellers swap from route 1 to route 2 on day $t + 1$ in response to the difference in route-costs on day t . If information is perfect then a naive adjustment process would see all drivers on route 1 swapping to route 2; this would plainly often oscillate and is, normally, unlikely to be realistic. On the other hand a high flow on the more costly route and a high cost difference, even if imperfectly perceived by travellers, would be likely to cause at least a few travellers to swap to route 2.

It is not clear how many travellers will swap for a given actual or perceived cost difference. Perhaps it is most simple to assume that the traveller flow swapping from route 1 to route 2 on day $t + 1$ is an increasing function of both

- the flow $X_1(t)$ (in vehicles per minute) on the more expensive route 1 on day t and
- the difference $C_1(\mathbf{X}(t)) - C_2(\mathbf{X}(t))$ in route costs on day t .

More complicated or general assumptions are possible; but this seems a reasonable place to start. Here $\mathbf{X}(t) = [X_1(t), X_2(t)]$ is the route flow vector on day t and $\mathbf{C}(\mathbf{X}(t))$ is the route-cost vector on day t .

Even if we stick to this simple two-route deterministic model, and follow the above assumptions, there is still a wide uncertainty concerning the number of travellers swapping from route 1 on day $t + 1$. In other words the rate of adjustment in the day to day assignment model is still, even in this really simple setup, something which is uncertain. Yet rates of adjustment are probably important if we wish to capture reality. There is much room for empirical studies.

Having made the above assumptions, perhaps the simplest swapping hypothesis is that the traveller flow swapping from route 1 to route 2 will be proportional to the product of the two factors above. If the only swaps are from higher to lower cost routes then for some constant $k > 0$, the changes $\Delta X_1(t), \Delta X_2(t)$ in the traveller flows on routes 1 and 2 will be given by the formulae:

$$\Delta X_1(t) = -kX_1(t) [C_1(\mathbf{X}(t)) - C_2(\mathbf{X}(t))] \tag{15a}$$

and

$$\Delta X_2(t) = +kX_1(t) [C_1(\mathbf{X}(t)) - C_2(\mathbf{X}(t))] \tag{15b}$$

These formulae depend on knowing that $C_1 - C_2 > 0$. To make equations (15) independent of this knowledge it is reasonable to define

$$x_+ = \max\{x, 0\}$$

for each real number x and then to extend equations (15) as follows:

$$\Delta X_1(t) = -kX_1(t) [C_1(\mathbf{X}(t)) - C_2(\mathbf{X}(t))]_+ + kX_2(t) [C_2(\mathbf{X}(t)) - C_1(\mathbf{X}(t))]_+ \tag{16a}$$

and

$$\Delta X_2(t) = -kX_2(t) [C_2(\mathbf{X}(t)) - C_1(\mathbf{X}(t))]_+ + kX_1(t) [C_1(\mathbf{X}(t)) - C_2(\mathbf{X}(t))]_+ \tag{16b}$$

This supposes that travellers on the cheaper route remain on that route. Given these rates (16a) and (16b) at which route 1 and route 2 flows change, the idealized day to day dynamical system becomes:

$$X_1(t + 1) = X_1(t) + \Delta X_1(t) \text{ and } X_2(t + 1) = X_2(t) + \Delta X_2(t)$$

or

$$\mathbf{X}(t + 1) = \mathbf{X}(t) + \Delta \mathbf{X}(t). \tag{17}$$

(16) ensures that

$$\Delta X_1(t) + \Delta X_2(t) = 0$$

so that if $X_1(0) + X_2(0) = T$ then $X_1(t) + X_2(t) = T$ for $t = 0, 1, 2, 3, \dots$. Total origin-destination flow is conserved by the route-swapping dynamical system. Moreover if $\mathbf{C}(\cdot)$ is continuous and k is small (and positive) then non-negativity is also conserved. [The agreement that $x_+ = \max\{x, 0\}$ for each real number x applies throughout this paper.]

Since t occurs above only as an argument in $\mathbf{X}(t)$ and $\Delta \mathbf{X}(t)$ there is another way of writing (16) and (17) which will prove useful when we come to generalise the dynamical system (16), (17). Let

$$\Delta_{12} = [-1, 1] = (-1, 1)^T \text{ and } \Delta_{21} = [1, -1] = (1, -1)^T.$$

Δ_{12} is the swap from route 1 to route 2 vector and Δ_{21} is the swap from route 2 to route 1 vector. Using these swap vectors we may define $\Delta \mathbf{X}$ by:

$$\Delta \mathbf{X} = k\{X_1[C_1(\mathbf{X}) - C_2(\mathbf{X})]_+ \Delta_{12} + X_2[C_2(\mathbf{X}) - C_1(\mathbf{X})]_+ \Delta_{21}\}, \tag{18}$$

and this equation may be written

$$\Delta \mathbf{X} = k \sum_{(r,s)} X_r [C_r(\mathbf{X}) - C_s(\mathbf{X})]_+ \Delta_{rs} \tag{19}$$

where the sum (19) is over the two ordered pairs (r, s) , following (18). (19) is a short ways of writing (16) (without t). We may include t again (recovering the dynamical system (17)) by putting:

$$\mathbf{X}(t + 1) = \mathbf{X}(t) + \Delta \mathbf{X}(t) \text{ for } t = 0, 1, 2, 3, \dots \tag{20}$$

To be slightly more general, assume given a fixed demand model with one OD pair joined by N routes, and suppose that the total OD flow is T (fixed). Then the dynamical system (19) and (20) still makes sense; only now there are more (r, s) pairs in the sum (19). In fact as there are N routes there are $N(N-1)$ pairs of routes and so the sum in (19) is in this case over $N(N-1)$ pairs (r, s) . It still holds that:

$$\sum_r \Delta X_r(t) = 0$$

for all t . Hence

$$\sum_r X_r(t+1) = \sum_r X_r(t) = \sum_r X_r(0) \text{ for all } t = 0, 1, 2, 3, \dots$$

So if the dynamical system (19), (20) starts at $\mathbf{X}(0) = \mathbf{X}^0$ within the set

$$F_1 = \{\mathbf{X}: \sum_r X_r = T\}$$

then it remains within that set. The feasible set must be further restricted to include only non-negative flows. So it is natural to put:

$$F = \{\mathbf{X}: \sum_r X_r = T, X_s \geq 0\}.$$

Now $\Delta\mathbf{X}$ (in (19)) is a continuous function of (current) flows and costs and so, for any given continuous cost flow function $\mathbf{C} = \mathbf{C}(\mathbf{X})$ defined on the whole feasible set F , $\Delta\mathbf{X}$ also becomes a continuous function of the current flow vector \mathbf{X} . It may then readily be shown that, if cost functions are continuous,

$$\mathbf{X} \geq \mathbf{0} \Rightarrow \mathbf{X} + \Delta\mathbf{X} \geq \mathbf{0}$$

provided k in (19) is sufficiently small (and positive). Thus if k is small enough and $\mathbf{X}(t)$ is feasible (belongs to F) then $\mathbf{X}(t) + \Delta\mathbf{X}(t)$ is also feasible. It follows that

$$\mathbf{X}(0) \in F \Rightarrow \mathbf{X}(t) \in F \text{ for all } t = 0, 1, 2, 3, \dots$$

Generalising further, suppose that there are many OD pairs but still N routes in total. It is now necessary to restrict swapping in (19) to only pairs of routes joining the same OD pair. One way of doing this is as follows. Specify the co-ordinates Δ_{rsq} of the N -vector Δ_{rs} as follows:

$$\Delta_{rsq} = \begin{cases} -1 & \text{if routes } r \text{ and } s \text{ join the same OD pair and } q = r, \\ +1 & \text{if routes } r \text{ and } s \text{ join the same OD pair and } q = s, \text{ and} \\ 0 & \text{otherwise.} \end{cases}$$

With this notation the dynamical system (19) and (20) remains valid. (Δ_{rs} is still the “swap from route r to route s vector”, and unsuitable swaps are ruled out.)

3.1 An anomaly and an initial solution

Consider figure 4 where link flows and costs are marked. The link flows must satisfy $x_1 + x_2 = x_3 + x_4 = T$ (and non-negativity constraints); such a link flow is called *feasible*. As has been pointed out recently by He et al (2010) the above dynamical system (19) and (20) is an unnatural day to day adjustment for this network. To see this, following He et al, suppose that link 1 and link 2 carry flow and that the node 2 \rightarrow node 1 traffic is equilibrated; then $c_1 = c_2$. Suppose also that the node 1 \rightarrow node 0 traffic is not equilibrated. If swap rule (19) and (20) is obeyed *for whole routes* then the “equilibrium” of the two left hand node 2 \rightarrow node 1 sub-routes will certainly be disturbed and some travellers will typically take longer to reach node 1 than they did before their swap. If we assume that travellers know that they would lose out by switching from the initial part of their current path (which reaches as far as node 1) then dynamical system (19) and (20) is wrong for this network. He et al (2010) propose a link flow adjustment process to avoid this anomaly.

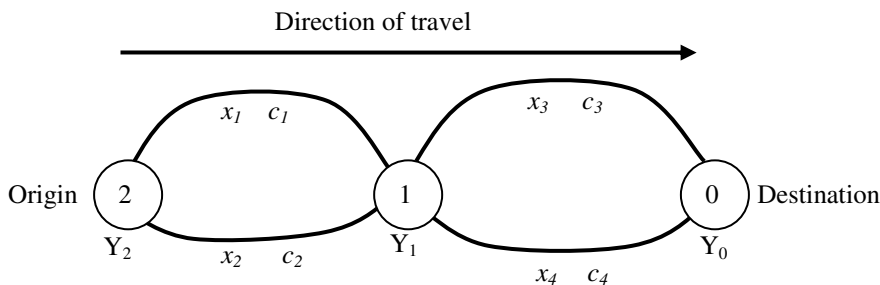


Figure 4. A four route network

In the rather different context of equilibrium-calculating procedures, Kupiszewska and Van Vliet (1999) considered changing the above dynamical system (19) and (20) by allowing only swaps toward the cheapest route; thus in (19) the sum then becomes a sum over just r (and s is fixed as the suffix of the cheapest route). Although this change was essentially aimed at making the computation of equilibria easier, this change does appear to remove the

anomaly identified by He et al. (2010) and so may be considered as a possible day to day dynamical system. (However this dynamical model is discontinuous when the least cost route changes, which is unrealistic.)

Nonetheless the following question arises: *can the dynamical system (19) and (20) be essentially retained but modified to deal with this day to day unrealism identified by He et al.?* The behavioural motivation here is that it may often be the case that drivers become aware of *some* cheaper route, but are unlikely to be always aware of *exactly* the cheapest route at all times. Further, it would seem intelligent to guide drivers toward a spectrum of cheaper routes rather than focussing traffic flows toward the one route which is currently the cheapest.

There are four main modifications which come to mind; all motivated by the network in figure 4. These are to apply (19) and (20) not to whole routes but to flows on bushes (Nie, 2010), to flows on paired alternative segments (Bar Gera (2009)), to approach proportions in the origin-based assignment method (Bar Gera (2002) and Bar-Gera and Boyce (2003)) or to splitting rates or flows exiting nodes (Gentile (2009) and Gentile and Knoedel (2009)). These possible modifications of (19) and (20) are likely to be closely related, and this relationship merits theoretical and computational investigation. In this paper we just focus on the splitting rate modification. This paper was motivated by considering the above work. Parts of the analysis presented here has close connections to the equilibrium-seeking methods presented by Bar Gera (2002). Appendix B gives a more detailed routing context for this paper.

Seeking to utilize splitting rates, and still considering the above example network in figure 4, assume now that drivers break their node-exit decision-making into two parts. One decision is taken at the origin (node 2) and involves deciding between just exit links 1 and 2, and a second decision is taken at node 1 and involves deciding between just exit links 3 and 4. This is similar in spirit to internet routing protocols. *Perhaps dynamical system (19), (20) may be applied to both decisions using the splitting rates at nodes instead of routes joining OD pairs?* In each of these decisions route costs to the destination via the different exit links available should perhaps inform each decision, replacing the whole route costs in (19) and (20) described above.

4. Splitting rate dynamics

The link flows in figure 4 may be thought of as node-exit flows, i.e. out of nodes 2 and 1. We then apply the day to day adjustment rules (19) and (20) to the node-exit flows x_1 and x_2 , and also independently to the node-exit flows x_3 and x_4 , but using flow-weighted costs to destinations instead of whole route costs. This will avoid the phenomenon highlighted by He et al (2010) and preserve the dynamical equations (19) and (20). Other issues (such as the generation of loops) may arise in general, but here we are making assumptions so that we may avoid these further issues; postponing their consideration. There are connections to the idea behind the Bellman-Ford algorithm (see Bertsekas and Gallager (1992), p. 396) for calculating shortest paths.

In the network of figure 4, suppose given a feasible vector of exit link flows, $\mathbf{x} = [x_1, x_2, x_3, x_4]$. The i^{th} co-ordinate x_i is to be regarded as a flow on link i and also as a flow to the destination via exit-link i . Since \mathbf{x} is feasible and so is non-negative and meets the OD demand $T > 0$,

$$x_1 + x_2 = x_3 + x_4 > 0, x_1 \geq 0, x_2 \geq 0, x_3 \geq 0, x_4 \geq 0.$$

The dynamical swap procedure is as follows. Specify in turn:

$$p_1 = \frac{x_1}{x_1 + x_2}, p_2 = \frac{x_2}{x_1 + x_2}, p_3 = \frac{x_3}{x_3 + x_4} \text{ and } p_4 = \frac{x_4}{x_3 + x_4};$$

$$Y_0 = 0;$$

$$Y_1 = p_3[c_3(x_3) + Y_0] + p_4[c_4(x_4) + Y_0];$$

and

$$Y_2 = p_1[c_1(x_1) + Y_1] + p_2[c_2(x_2) + Y_1].$$

Here (p_1, p_2) are the splitting rates at node 2; (p_3, p_4) are the splitting rates at node 1; Y_1 is the flow-weighted cost to the destination from node 1 (minutes per vehicle); and Y_2 is the flow-weighted cost to the destination from node 2. ($Y_0 = 0$ is the flow-weighted cost to the destination node 0 from node 0 itself.) The dependence of the Y_i on \mathbf{x} is omitted. Given \mathbf{x} and the Y_i , the flow-weighted costs of travel to the destination via all the links 1, 2, 3, and 4 are then:

$$C_4(\mathbf{x}) = c_4(x_4) + Y_0, C_3(\mathbf{x}) = c_3(x_3) + Y_0, C_2(\mathbf{x}) = c_2(x_2) + Y_1, \text{ and } C_1(\mathbf{x}) = c_1(x_1) + Y_1.$$

These flow-weighted costs are now used to swap the flows exiting nodes 1 and 2 following formulae similar to (19) and (20) above, as follows. Changes in outflow vectors from nodes 1 and 2 are determined separately, using the above flow-weighted costs and also “swap” vectors. We let

$$\Delta^1 \mathbf{x} = k \{x_3[C_3(\mathbf{x}) - C_4(\mathbf{x})]_+ \Delta_{34} + x_4[C_4(\mathbf{x}) - C_3(\mathbf{x})]_+ \Delta_{43}\}; \text{ and}$$

$$\Delta^2 \mathbf{x} = k \{x_1[C_1(\mathbf{x}) - C_2(\mathbf{x})]_+ \Delta_{12} + x_2[C_2(\mathbf{x}) - C_1(\mathbf{x})]_+ \Delta_{21}\}.$$

Here $\Delta_{34q} = -1$ if $q = 3$, $+1$ if $q = 4$ and 0 otherwise and similarly $\Delta_{12q} = -1$ if $q = 1$, $+1$ if $q = 2$ and 0 otherwise. With these definitions, $\Delta^1 \mathbf{x}$ swaps node-exit flow just at node 1 and $\Delta^2 \mathbf{x}$ swaps node-exit flow at just node 2. The total $\Delta \mathbf{x} = \Delta^1 \mathbf{x} + \Delta^2 \mathbf{x}$ and so (rather like (19)):

$$\Delta \mathbf{x} = k \sum_{(i,j)} x_i [C_i(\mathbf{x}) - C_j(\mathbf{x})]_+ \Delta_{ij}. \tag{21}$$

The sum in (21) takes account of the restriction that swaps only occur between exit links at a single node by the definitions of Δ_{34} and Δ_{12} . So the only non-zero terms in the sum (21) are those corresponding to those $(i, j) = (1, 2), (2, 1), (3, 4)$ and $(4,3)$. The day to day evolution of the system is then to be, for any given feasible starting link-flow vector \mathbf{x}^0 :

$$\mathbf{x}(0) = \mathbf{x}^0; \text{ and} \tag{22}$$

$$\mathbf{x}(t + 1) = \mathbf{x}(t) + \Delta \mathbf{x}(t) \text{ for } t = 0, 1, 2, 3, \dots$$

(21), (22) is similar to the route-swapping dynamical system (19), (20) but uses different variables (node exit flows or splitting rates instead of route-flows). It is clear that now, under this link-based or node-exit-based dynamical system, if the two partial paths from node 2 to node 1 have equal cost; that is $c_1 = c_2$; then the flows x_1, x_2 from node 2 to node 1 will not change and the behavioural anomaly identified by He et al has been removed. It is easy to check in this example network that (21) preserves conservation of flow at nodes; and that \mathbf{x} is a Wardrop equilibrium link flow if and only if $\Delta \mathbf{x} = \mathbf{0}$.

$\Delta \mathbf{x}$ (in (21)) is a continuous function of flows and costs and so, for any given continuous cost flow function $\mathbf{c}(\mathbf{x})$, $\Delta \mathbf{x}/k$ becomes (as previously) a continuous function of today’s link flow vector. It follows that,

$$\mathbf{x} \geq \mathbf{0} \Rightarrow \mathbf{x} + \Delta \mathbf{x} \geq \mathbf{0}$$

provided k is sufficiently small. This implies that if the assignment \mathbf{x} is feasible (a non-negative link flow vector consistent with the fixed total OD flow T) then $\mathbf{x} + \Delta \mathbf{x}$ is also feasible. We consider the dynamical system (22) with this specification of $\Delta \mathbf{x}$ in (21), where k is small. More extensive single-destination networks may be allowed for as follows. Let the total number of links be N and allow the sum in (21) to be over all $N(N-1)$ pairs (i, j) but specify the N -vector Δ_{ij} as follows:

$$\Delta_{ijq} = \begin{cases} -1 & \text{if links } i \text{ and } j \text{ lead away from the same node and } q = i, \\ +1 & \text{if links } i \text{ and } j \text{ lead away from the same node and } q = j, \text{ and} \\ 0 & \text{otherwise.} \end{cases}$$

Again, with this notation unsuitable swaps are ruled out with no explicit constraint on the pairs (i, j) in the sum (21).

4.1 Extending the splitting rate model above to a slightly more complicated network

Consider the network in figure 5 where again a feasible link flow vector $\mathbf{x} = [x_1, x_2, x_3, x_4, x_5]$ is marked. In this case we need to ensure that we take account of changes in x_5 when changing the flows x_1, x_2, x_3 and x_4 , to ensure that flows are still conserved at nodes 2 and 1. We must now have:

$$x_1 + x_2 + x_5 = T$$

and

$$x_1 + x_2 = x_3 + x_4$$

(and non-negativity constraints).

Does the above adjustment still work? As before, let \mathbf{x} be a feasible link flow vector and calculate in order:

$$p_1 = \frac{x_1}{x_1 + x_2 + x_5}, p_2 = \frac{x_2}{x_1 + x_2 + x_5}, p_5 = \frac{x_5}{x_1 + x_2 + x_5}, p_3 = \frac{x_3}{x_3 + x_4} \text{ and } p_4 = \frac{x_4}{x_3 + x_4};$$

$$Y_0 = 0;$$

$$Y_1 = p_3[c_3(x) + Y_0] + p_4[c_4(x) + Y_0]; \text{ and}$$

$$Y_2 = p_1[c_1(x) + Y_1] + p_2[c_2(x) + Y_1] + p_5[c_5(x) + Y_0].$$

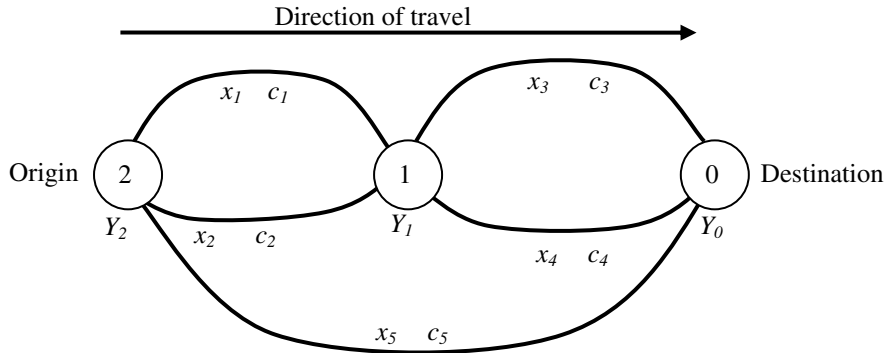


Figure 5. A five route network

(We assume that $x_3 + x_4 > 0$.) Again using these \mathbf{x} -dependent Y values let:

$$C_5(\mathbf{x}) = c_5(x_5) + Y_0; C_4(\mathbf{x}) = c_4(x_4) + Y_0; C_3(\mathbf{x}) = c_3(x_3) + Y_0; C_2(\mathbf{x}) = c_2(x_2) + Y_1; \text{ and } C_1(\mathbf{x}) = c_1(x_1) + Y_1.$$

Using these flow-weighted costs to the destination, the update directions for the (link-flow or) node-outflow vector \mathbf{x} are now to be:

$$\delta^1 \mathbf{x} = k \{ x_3 [C_3(\mathbf{x}) - C_4(\mathbf{x})]_+ \Delta_{34} + x_4 [C_4(\mathbf{x}) - C_3(\mathbf{x})]_+ \Delta_{43} \}$$

at node 1 and

$$\delta^2 \mathbf{x} = k \{ x_1 [C_1(\mathbf{x}) - C_2(\mathbf{x})]_+ \Delta_{12} + x_2 [C_2(\mathbf{x}) - C_1(\mathbf{x})]_+ \Delta_{21} + x_5 [C_5(\mathbf{x}) - C_1(\mathbf{x})]_+ \Delta_{51} + x_1 [C_1(\mathbf{x}) - C_5(\mathbf{x})]_+ \Delta_{15} + x_5 [C_5(\mathbf{x}) - C_2(\mathbf{x})]_+ \Delta_{52} + x_2 [C_2(\mathbf{x}) - C_5(\mathbf{x})]_+ \Delta_{25} \}$$

at node 2.

The total or whole adjustment vector $\delta \mathbf{x} = \delta^1 \mathbf{x} + \delta^2 \mathbf{x}$ may now be written in vector form as before:

$$\delta \mathbf{x} = k \sum_{(i,j)} x_i [C_i(\mathbf{x}) - C_j(\mathbf{x})]_+ \Delta_{ij}. \tag{23}$$

The sum is essentially over just eight ordered pairs (i, j) : (3, 4), (4, 3) arising at node 1 and (1, 2), (2, 1), (1, 5), (5, 1), (2, 5) and (5, 2) arising at node 2. Thus again the direction (23) is a sum of swaps, each between a single pair of link exit flows at a single node. Suppose that k is small enough to ensure that non-negative link flows in \mathbf{x} remain non-negative in $\mathbf{x} + \delta \mathbf{x}$.

Direction (23) (like (21)) is now in exactly the same form as (19) above and so is an exit-link version of the dynamical suggestion in Smith (1984a). But now we only obtain a provisional flow vector \mathbf{x}^{prov} since conservation at node 1 is lost:

$$\mathbf{x}^{\text{prov}} = \mathbf{x} + \delta \mathbf{x} \tag{24}$$

where the superfix prov stands for “provisional”.

Network loading is now required to recover flow conservation following the new splitting rates. One way of doing this is to use an iterative process as follows. At iteration u , suppose that the provisional link flow vector is $\mathbf{x}^{\text{prov}}(u)$. For those nodes where conservation fails (just node 1 here) multiply all the provisional outflows from such a node so as to match (in total) the provisional inflows at that node. The provisional link flow vector $\mathbf{x}^{\text{prov}}(u)$ will be modified and becomes $\mathbf{x}^{\text{prov}}(u + 1)$. Provided the network is loop-free, after several iterations of this modification conservation at all nodes will be restored. At this point we shall say that the flows are no longer provisional but “actual”. To be more explicit; in iteration u , at each node n where conservation does not hold we put (for each link i which is outgoing at n):

$$x_i^{\text{prov}}(u + 1) = \{ [\sum_r \text{an incoming link at } n x_r^{\text{prov}}(u)] / [\sum_r \text{an outgoing link at } n x_r^{\text{prov}}(u)] \} x_i^{\text{prov}}(u). \tag{25}$$

[Here, in general, the flow generated at origin nodes at the entrances of links r must be properly accounted for in the numerators in the “balancing” iteration u given by (25); if for example in figure 5 flow was also generated at node 1 then this would be included in the numerator of the update equation (25) corresponding to node 0.] The whole; (23), (24) and (25); may be regarded as an exit-link form of the route adjustment algorithm (19) and (20). The result of these adjustments (25) is the “true” change $\Delta \mathbf{x}$ in the flow vector \mathbf{x} , arising from the provisional change $\delta \mathbf{x}$. With many destinations, a copy of this modification procedure is needed for each destination and then the procedure is

applied to each copy. (It might be possible that iterations (25) may be done just once for each node; if the nodes can be suitable ordered.)

4.2 Properties of the above link flow adjustment for a general one destination network

In order to generalise we make the following three assumptions which apply to all networks in the remainder of the paper.

1. The network contains no loops so that any link flow vector \mathbf{x} is loop-free.
2. The network contains a single destination node.
3. There is a fixed flow rate from each non-destination node to the single destination node.

Assume that \mathbf{x} is a feasible (loop-free) link flow vector. Since the total flow out of each non-destination node is positive, the co-ordinates of this vector determine the splitting rate vector \mathbf{p} (with co-ordinates $p_i = p(i)$): for each link i , $p(i)$ is the proportion of the total flow passing through that node n upstream of i which exits node n via link i . For this reason link i may always be thought of as exit link i ; it is a link allowing exit from the headnode of the link. Since there is outflow from every non-destination node exit proportions are defined at every non-destination node.

First we consider the effects of downstream link costs on upstream flow-weighted costs to the destination. Suppose that there is a link (or exit-link) cost-flow function $\mathbf{c} = \mathbf{c}(\mathbf{x})$, determining the exit-link cost vector $\mathbf{c}(\mathbf{x})$ in terms of the exit-link flow vector \mathbf{x} . Let C_i denote the (flow-weighted average) cost to the single destination via link i . This is determined by the vector $\mathbf{c}(\mathbf{x})$ of link costs and the vector \mathbf{p} of splitting rates and so depends on the link flow vector \mathbf{x} . The flow-weighted cost-to-destination vector \mathbf{C} (with one co-ordinate for each link) is given by:

$$\mathbf{C} = \mathbf{C}(\mathbf{x}) = \mathbf{M}\mathbf{c}(\mathbf{x}) \tag{26}$$

where $M_{ii} = 1$ for all i and each off-diagonal entry of \mathbf{M} is a sum of products of components of the vector \mathbf{p} . To be precise, whenever $j \neq i$:

$$M_{ij} = \sum p(i_1)p(i_2)\dots p(i_{m-1})p(j)$$

where the sum is over all sequences of links i_1, i_2, \dots, i_{m-1} , which are contiguous and lead from the downstream node of link i to the upstream node of link j (where $j \neq i$). Each such product is the proportion of the flow along link i which utilises that particular subroute leading from link i to link j . The sum of such products is the proportion of flow through link i which also passes through link j , and so experiences the cost on link j , en-route to the single destination. To obtain the cost to the destination from link i we must therefore multiply this proportion M_{ij} by $c_j(\mathbf{x})$ and add over j as suggested in (26).

Now, second, we consider the effects of upstream flow changes on downstream link flows. Let δx_i be a provisional change in flow on link i (given by (23)) and consider the effect of this single change on the flow on a link j downstream from link i . The change is brought about by using (25), as well as (23). A little thought shows that the consequential flow change on link j will be

$$\sum p(i_1)p(i_2)\dots p(i_{m-1})p(j) \times \delta x_i$$

where the sum is over all sequences of directed links $i_1, i_2, \dots, i_j, \dots, i_{m-1}$, which are contiguous and lead from the downstream node of link i to the upstream node of link j ($\neq i$).

It is now natural to add the effects on link- j -flow of all upstream changes such as δx_i . To do this we let

$$N_{ji} = \sum p(i_1)p(i_2)\dots p(i_{m-1})p(j) \tag{27}$$

where the sum is over all sequences of links $i_1, i_2, \dots, i_j, \dots, i_{m-1}$, which are contiguous and lead from the downstream node of link i to the upstream node of link j ($\neq i$). Each product in this sum is the proportion of that flow change δx_i implemented on link i felt (via updates (25)) by the flow along that subroute leading from link i to link j . Then the sum of all changes on link j , due to flow changes δx_i on all upstream links i , is

$$\sum_i N_{ji} \delta x_i.$$

It follows that the change $\Delta \mathbf{x}$ in the whole flow vector \mathbf{x} deriving from the whole provisional change vector $\delta \mathbf{x}$ is given by:

$$\Delta \mathbf{x} = \mathbf{N}\delta \mathbf{x}$$

where $N_{jj} = 1$ for all j and, for $i \neq j$, N_{ji} is given by (27). (Each off-diagonal entry of \mathbf{N} is, as with \mathbf{M} above, a sum of products of components of \mathbf{p} .)

It is thus clear from the above equations that $\mathbf{N} = \mathbf{M}^T$ so we have both:

$$\mathbf{C}(\mathbf{x}) = \mathbf{M}\mathbf{c}(\mathbf{x}) \text{ and } \Delta\mathbf{x} = \mathbf{M}^T\delta\mathbf{x}. \tag{28}$$

Here \mathbf{M} depends on \mathbf{p} and so on \mathbf{x} .

4.3 Objective function

Let D be the set of demand-feasible link flow vectors and let $\mathbf{c}(\cdot)$ determine link costs in terms of the vector \mathbf{x} of link flows. Suppose that $\mathbf{c}(\mathbf{x})$ is defined for all \mathbf{x} in D . Then the link flow vector \mathbf{x}^* (in D) is a Wardrop equilibrium link flow if and only if

$$-\mathbf{c}(\mathbf{x}^*) \text{ is normal at } \mathbf{x}^* \text{ to } D. \tag{29}$$

See Smith (1979).

Suppose now that each link cost function c_i is a continuous function of just the flow x_i along link i alone and let

$$Z(\mathbf{x}) = \sum_i \int_0^{x_i} c_i(u) du = \sum_i Z_i(x_i). \tag{30}$$

Then Z is the standard (inelastic) Beckmann objective function (Beckmann et al (1956)). As each link cost c_i depends only on the flow x_i along link i :

$$\frac{\partial Z}{\partial x_i} = c_i(x_i)$$

and so

$$\text{grad } Z(\mathbf{x}) = \mathbf{c}(\mathbf{x}) \tag{31}$$

Thus \mathbf{x}^* is an equilibrium if and only if

$$-\text{grad } Z(\mathbf{x}^*) = -\mathbf{c}(\mathbf{x}^*) \text{ is normal at } \mathbf{x}^* \text{ to } D. \tag{32}$$

It follows from (29) - (32) that if \mathbf{x} is feasible but not an equilibrium then

$$-\mathbf{c}(\mathbf{x}) \text{ is not normal at } \mathbf{x} \text{ to } D$$

and hence that there is a direction δ from \mathbf{x} such that $\mathbf{x} + \delta \in D$ and

$$\delta \cdot [-\text{grad } Z(\mathbf{x})] = \delta \cdot [-\mathbf{c}(\mathbf{x})] > 0.$$

This direction δ is thus a feasible Z -descent direction at \mathbf{x} .

Let \mathbf{x} be a non-equilibrium. Then by above there is a feasible Z -descent direction in D at \mathbf{x} . It is natural to move in such directions aiming to approach the set of \mathbf{x}^* satisfying (32), that is, the set of equilibria. Many equilibration algorithms follow this line. Here however we are only looking at the special direction $\Delta\mathbf{x}$ in (28), generated by the specific direction $\delta\mathbf{x}$ using (23), (24) and (25). *Is $\Delta\mathbf{x}$ a Z -descent direction ?*

4.4 Outline proof of descent at a non-equilibrium flow vector \mathbf{x} , using $\Delta\mathbf{x}$

The network is loop-free and each node has flow headed for the single destination. Throughout this section \mathbf{x} is to be an arbitrary non-equilibrium link flow vector in D . We consider the direction $\Delta\mathbf{x}$ generated by $\delta\mathbf{x}$ following (23), (24), and (25) or (28). This direction $\Delta\mathbf{x}$ gives rise to the splitting rate re-routeing model.

Suppose that (non-equilibrium) link flow vector \mathbf{x} is changed provisionally by adding $\delta\mathbf{x}$, where $\delta\mathbf{x}$ is given by (23). Then it follows directly from the definition of $\delta\mathbf{x}$ that

$$\delta\mathbf{x} \neq \mathbf{0} \Rightarrow (\delta\mathbf{x})^T \mathbf{C}(\mathbf{x}) < 0. \tag{33}$$

Of course $\mathbf{x} + \delta\mathbf{x}$ will not necessarily be feasible as conservation will typically be lost. (In which case $\mathbf{x} + \delta\mathbf{x}$ will be provisional.) But (24) and (25), or (28), may then be utilised to give a corresponding actual feasible change in the link flow vector \mathbf{x} of $\Delta\mathbf{x} = \mathbf{M}^T\delta\mathbf{x}$. Then, using (28) twice and (33):

$$\Delta\mathbf{x}^T \text{grad } Z(\mathbf{x}) = \Delta\mathbf{x}^T \mathbf{c}(\mathbf{x}) = (\mathbf{M}^T\delta\mathbf{x})^T \mathbf{c}(\mathbf{x}) = (\delta\mathbf{x})^T \mathbf{M}\mathbf{c}(\mathbf{x}) = (\delta\mathbf{x})^T \mathbf{C}(\mathbf{x}) < 0 \tag{34}$$

provided $\delta\mathbf{x} \neq \mathbf{0}$. Thus, to ensure that $\Delta\mathbf{x} = \mathbf{M}^T\delta\mathbf{x}$ is a Z -descent direction at the non-equilibrium \mathbf{x} in D , we must just show that $\delta\mathbf{x} \neq \mathbf{0}$ at the non-equilibrium \mathbf{x} .

Now at each node n let the truly least cost to the destination from n be $L_n(\mathbf{x})$. We also have the flow-weighted cost, $Y_n(\mathbf{x})$, to the destination from n . Thirdly, for each link i leaving n we have the flow weighted cost $C_i(\mathbf{x})$ to the destination via link i defined as follows:

$$C_i(\mathbf{x}) = c_i(\mathbf{x}) + Y_m(\mathbf{x})$$

where m is the end node of link i .

Given the non-equilibrium \mathbf{x} , let $S(\mathbf{x})$ be the set of nodes such that $L_n(\mathbf{x}) = Y_n(\mathbf{x})$. We now sometimes omit the argument \mathbf{x} . The set of nodes in S is non-empty as the destination node, node 0, is in S . Also the set of nodes not in S is non-empty. Because if it were empty then $L_n = Y_n$ for all n and all flow must be leaving n along a shortest path for all nodes n . In this case no flow is on a more expensive route and we are at a Wardrop equilibrium.

Since \mathbf{x} is not an equilibrium, it follows that there must be a link $j = (n, m)$ along a shortest path with n not in S and m in S . To see this, start at any node not in S and follow a shortest path to the destination until a node in S is first reached. Call this m . Backtrack one node from m to obtain node n , the predecessor of m along the link (say link j) just traversed. Link $j = (n, m)$ lies along a shortest path from n (not in S) to m (in S).

By construction the above link j is on a shortest path from $n \notin S(\mathbf{x})$, and also the exit node of link j , namely m , belongs to $S(\mathbf{x})$. The argument below uses both these facts. Since link j is on a shortest path from n :

$$L_n(\mathbf{x}) = c_j(\mathbf{x}) + L_m(\mathbf{x})$$

and since $m \in S$

$$L_m(\mathbf{x}) = Y_m(\mathbf{x}).$$

Hence

$$L_n(\mathbf{x}) = c_j(\mathbf{x}) + L_m(\mathbf{x}) = c_j(\mathbf{x}) + Y_m(\mathbf{x}) = C_j(\mathbf{x}) \tag{35}$$

the flow-weighted average cost to the destination via link j .

Since all non-destination nodes generate positive exiting flow and since n is by construction not in S there must be some flow which leaves node n via an exit link which does not lie along a least cost route to the destination. Suppose that node- n -exit link i has positive flow, and is not on a least cost route to the destination from n . Then:

$$x_i > 0 \text{ and } C_i(\mathbf{x}) > L_n(\mathbf{x}).$$

However $L_n(\mathbf{x}) = C_j(\mathbf{x})$ by (35); hence

$$x_i > 0 \text{ and } C_i(\mathbf{x}) > C_j(\mathbf{x})$$

and so

$$x_i [C_i(\mathbf{x}) - C_j(\mathbf{x})] > 0.$$

It follows that

$$x_i [C_i(\mathbf{x}) - C_j(\mathbf{x})] \Delta_{ij} \neq 0$$

since links i and j are both exiting node the same node (node n). It follows (now allowing i and j to be general suffices) that

$$\delta \mathbf{x} = \sum_{(i,j)} x_i [C_i(\mathbf{x}) - C_j(\mathbf{x})] \Delta_{ij} \neq \mathbf{0}.$$

Cancellation is impossible by virtue of the construction of the vectors $x_i [C_i(\mathbf{x}) - C_j(\mathbf{x})] \Delta_{ij}$; if one term of the sum is non-zero so is the whole sum.

Since $\delta \mathbf{x} \neq \mathbf{0}$, it now follows from (33) and (34) that

$$(\Delta \mathbf{x})^T \text{grad } Z(\mathbf{x}) = (\delta \mathbf{x})^T \mathbf{C}(\mathbf{x}) < 0.$$

\mathbf{x} was here an arbitrary non-equilibrium flow vector and so $\Delta \mathbf{x}$ is a feasible descent direction for Z at \mathbf{x} , for any non-equilibrium flow vector \mathbf{x} .

5. Extending the above theory to include responsive signal control

Traffic assignment and traffic control have been considered together previously by, for example, Van Vuren and van Vliet (1992). For simplicity here we will still consider the single destination framework in section 4 including the assumptions 1 – 3 given in section 4; however the analysis here may be repeated on a multi-destination network using a multi-level representation, following Charnes and Cooper (1958) and Aashtiani and Magnanti (1983).

A splitting rate model would seem extremely appropriate for incorporating signal green-times, since the way green-time is split among the various stages at a junction is central to any study of traffic control. Various signal control policies may be easily included within the above splitting rate framework as follows. At each node all the incoming links are divided into stages (maximal sets of compatible approaches) and all links in a stage are given green simultaneously. Suppose that stage K is green for a proportion of time G_K . Let antistage K be the set of all approaches *not* in stage K ; then approaches in antistage K are shown red simultaneously when stage K is shown green. Thus antistage K is red for a proportion of time R_K equal to G_K .

A simple way of adding control as a new variable within the splitting rate assignment model above is to think of the red time awarded to an antistage as a different type of flow taking up some of the available capacity at the exits of those approaches in that antistage. Thus the aggregated flow on approach i will comprise the flow of real vehicles added to a suitable multiple of “red-time” (designed to take up the capacity which cannot be used while the signal is red for that approach). For each link i we let the new “flow volume” $v_i = x_i + s_i r_i$; where x_i represents the “real” vehicular flow and r_i represents the proportion of time approach i is red. The multiple $s_i r_i$ (vehicles per minute) is the capacity lost due to the proportion (r_i) of red time, bearing in mind the saturation flow s_i (vehicles per minute) at the link exit. Then we suppose that the cost of traversing approach i equals

$$c_i(x_i) + b_i(x_i + s_i r_i).$$

Here $c_i(x_i)$ now represents the cost of traversing the length of the link when the flow is x_i and b_i represents the bottleneck delay felt at the traffic signal when the flow is x_i and the red time is r_i . Both $c_i(\cdot)$ and $b_i(\cdot)$ are non-decreasing functions. Here the slope of c_i may be rather flat and the slope of b_i may be rather steep: b_i may even have a vertical asymptote at s_i .

This approach allows a simple dynamical model of control and routing to be constructed. In essence we will now have a *two commodity* link model where the two commodities are:

- x_i = vehicular flow on link i (vehicles per minute) and
- r_i = red-time on link i (a proportion and dimensionless).

The dynamics are now to be as follows. At each node: real flow switches to cheaper exit-links using flow-weighted exit-link costs $C_i(\mathbf{x})$ generated as in previous sections by link costs $c_i(x_i) + b_i(x_i + r_i s_i)$. Also red-time switches to “cheaper” antistages using link “red-time costs” which are defined to suit the particular signal control policy employed. For example,

- for the equisaturation policy this “red-time cost” is $x_i/g_i s_i = x_i/[(1-r_i)s_i] = x_i/[s_i - s_i r_i]$ and
- for the P_0 policy (see Smith (1979)) this “red-time cost” is $s_i b_i(x_i + s_i r_i)$.

(Observe that both of the above red-time costs are increasing functions of r_i so both “red-time cost functions” seem initially sensible from an assignment viewpoint.)

It may be seen from the above swapping specifications that at a junction with two approaches the P_0 policy chooses red times which seek to ensure that

$$s_1 b_1(x_1 + s_1 r_1) = s_2 b_2(x_2 + s_2 r_2),$$

as if this holds then no swapping occurs in this P_0 case; and the equisaturation policy chooses green times which seek to ensure that

$$x_1/g_1 s_1 = x_2/g_2 s_2,$$

as if this holds then no swapping occurs in the equisaturation case. It is clear from the above equation in the P_0 case that if the saturation flow s_2 is high then the P_0 policy will (by a suitable choice of \mathbf{r}) seek to ensure that the bottleneck delay b_2 will tend to be small; encouraging the use of the approach with higher saturation flows (even if actual flows on that approach are small). Thus the policy encourages re-routing rather than rewarding travellers on existing routes. It may be shown that, under natural conditions, this policy maximises network throughput at an equilibrium distribution of traffic flows. (See Smith (1979).)

5.1 Outline of the extension to include the responsive P_0 control policy within the splitting rate framework

Here we extend, in outline only, our previous splitting rate descent / convergence / stability results in section 4 and in the appendix D so as to include P_0 within the day to day dynamic framework. Suppose then that the P_0 signal control policy is employed at each node of a network. Then the 2-commodity link cost-flow function is, following the previous section:

$$[c_i(x_i) + b_i(x_i + s_i r_i), s_i b_i(x_i + s_i r_i)].$$

Extend the standard Beckmann expression (30) to include signal timings, by defining

$$W(\mathbf{v}) = \sum_i \int_0^{v_i} b_i(u) du = \sum_i W_i(v_i)$$

and

$$V(\mathbf{x}, \mathbf{r}) = \sum_i (Z_i(x_i) + W_i(x_i + s_i r_i)).$$

Then:

$$\frac{\partial V}{\partial x_i} = \frac{\partial [Z_i(x_i) + W_i(x_i + s_i r_i)]}{\partial x_i} = c_i(x_i) + b_i(x_i + s_i r_i)$$

and

$$\frac{\partial V}{\partial r_i} = \frac{\partial [Z_i(x_i) + W_i(x_i + s_i r_i)]}{\partial r_i} = s_i b_i(x_i + s_i r_i).$$

It follows that

$$[\text{grad } V(\mathbf{x}, \mathbf{r})]_i = [c_i(x_i) + b_i(x_i + s_i r_i), s_i b_i(x_i + s_i r_i)]$$

and so

$$\text{grad } V(\mathbf{x}, \mathbf{r}) = [\mathbf{c}(\mathbf{x}) + \mathbf{b}(\mathbf{x} + \mathbf{s}\mathbf{r}), \mathbf{s}\bullet\mathbf{b}(\mathbf{x} + \mathbf{s}\mathbf{r})].$$

The vector $\mathbf{s}\bullet\mathbf{r}$ is called the Hadamard product of the vectors \mathbf{s} and \mathbf{r} ; it is the column vector with i^{th} co-ordinate $s_i r_i$. Also $\mathbf{s}\bullet\mathbf{b}(\mathbf{x} + \mathbf{s}\mathbf{r})$ is the vector with i^{th} co-ordinate $s_i b_i(x_i + s_i r_i)$. It follows at once as in the usual Beckman argument applied in section 4 above to just routing changes, that (away from equilibrium) switching flow toward cheaper node exits (and then applying (25)) and swapping red-time toward cheaper anti-stages, at each node, will decrease

$$V(\mathbf{x}, \mathbf{r}) = \sum_i (Z_i(x_i) + W_i(x_i + s_i r_i))$$

where the summation is over all links i . At a stationary point of V , at each node all exit flows are on cheapest link exits at that node and all red times are on “cheapest” antistages at that node. This means that Wardrop’s equilibrium condition holds and the signals satisfy the P_0 control policy, simultaneously.

Both the flow switches and the red-time switches above depend on the specifications of the costs of exit links and antistages at a node. For each node n :

- the *exit-link i cost*, C_i , used to switch exit flows toward cheaper exit links at node n , is the flow-weighted cost-to-destination in (26) in section 4.1 above, but with $c_i(x_i)$ replaced by $= c_i(x_i) + b_i(x_i + s_i r_i)$ and
- the *antistage J red cost*, $RC_J = RC_J(\mathbf{x}, \mathbf{r})$, used to switch red-time toward cheaper antistages at a node, is:

$$RC_J = \sum_{i \in A_J} s_i b_i(x_i + s_i r_i).$$

Here A_J is the set of links i in antistage J . The formula here for RC_J arises from combining the definition of the P_0 signal control policy in Smith (1979) and the simple delay formulation introduced in this section 5 above.

The combined (link flow, red-time) direction is thus $[\Delta\mathbf{x}, \Delta\mathbf{r}]$ where $\Delta\mathbf{x}$ arises via (24) and (25) or (28) from

$$\delta\mathbf{x} = \sum_{(i,j)} k [C_i - C_j]_+ x_i \Delta_{ij}$$

as previously, and $\Delta\mathbf{r}$ arises in a similar way via the stage-link incidence matrix from

$$\delta\mathbf{R} = \sum_{(I,J)} k [RC_I - RC_J]_+ R_I \Delta_{IJ}. \tag{36}$$

Here, for two antistages I and J at the same node, Δ_{IJ} has -1 in the I^{th} place and +1 in the J^{th} place and zeros everywhere else. Δ_{IJ} is the red-time swap vector from antistage I to antistage J at the same node. [Δ_{IJ} is the zero vector if anti-stages I and J belong to different nodes.]

Explicitly, let \mathbf{S} be the stage matrix: $S_{ji} = 1$ if link i is in stage J and 0 otherwise. Then the change $\Delta\mathbf{r}$ in the link red time vector \mathbf{r} arising from the change $\delta\mathbf{R}$ in (36) is given by:

$$\Delta\mathbf{r} = \mathbf{S}^T \delta\mathbf{R}$$

just as the change $\Delta\mathbf{x}$ in the link flow vector \mathbf{x} is given as in (28) above by:

$$\Delta\mathbf{x} = \mathbf{M}^T \delta\mathbf{x}.$$

As with just flow changes, $(\Delta\mathbf{x}, \Delta\mathbf{r})$ is a descent direction for $V(\mathbf{x}, \mathbf{r})$ (at any (\mathbf{x}, \mathbf{r}) which is not an equilibrium consistent with the P_0 policy). This may be shown by a simple enhancement of the proof of descent for just flow change $\Delta\mathbf{x}$ given in section 4.4 above. The combined flow, red-time dynamical system starts at any feasible start point $(\mathbf{x}^0, \mathbf{r}^0)$, and is:

$$(\mathbf{x}(0), \mathbf{r}(0)) = (\mathbf{x}^0, \mathbf{r}^0) \text{ and}$$

$$(\mathbf{x}(t+1), \mathbf{r}(t+1)) = (\mathbf{x}(t), \mathbf{r}(t)) + h(\Delta\mathbf{x}(t), \Delta\mathbf{r}(t)) \text{ for } t = 0, 1, 2, 3, 4 \dots$$

where $h > 0$. A modification of the proof in appendix D below shows that this dynamical system approaches the equilibrium set arbitrarily closely; if h is sufficiently small.

The (link flow, red-time) vector (\mathbf{x}, \mathbf{r}) is here being forced, by following $[\Delta\mathbf{x}, \Delta\mathbf{r}]$, to a set of approximately stationary points of V subject to:

- flow conservation and non-negativity constraints and*
- antistage red time conservation and non-negativity constraints*

at each node. It is assumed that these hold at the start; then using only the not-too-large switches specified in this section ensures that these constraints continue to hold.

6. Comparison between optimum (bilevel) timings and P_0 timings on a simple network

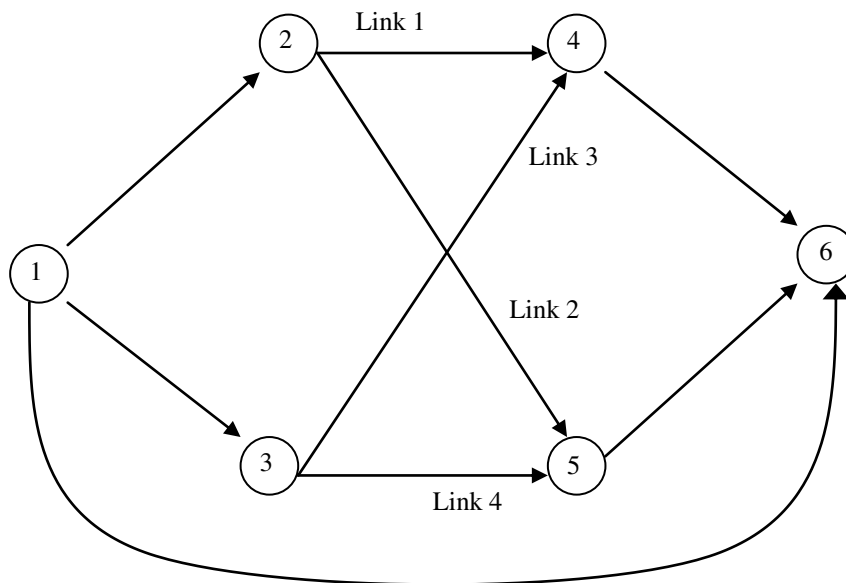


Figure 6. A simple capacitated signal controlled network

The network in figure 6 was first considered by Hai Yang (1996). The network has one origin (node 1), one destination (node 6) and two signals at nodes 4 and 5. All links have a finite saturation flow and delay on any link tends to infinity as the capacity of the link is approached. Thus the network is capacitated. Links 1, 2, 3 and 4 are signal-controlled. In Clegg et al (2001) two different signal timings were implemented at each of 328 steady demand levels:

- (a) the approximately optimum timings taking into account route choices, using bilevel optimisation, and
- (b) the timings which are part of a consistent equilibrium with the responsive P_0 policy.

For each demand level implementation (a) required a rather complicated bilevel optimisation. Implementation (b) was much simpler as the iterative calculational process depends much more simply on the flows and delays at the signal. The central results are shown in figures 7 and 8.

Results are plotted for each of 328 demand levels; a demand beyond level 328 is beyond the capacity of the network. The details of the bilevel method and the network are given in Clegg et al (2001) and Smith (2006).

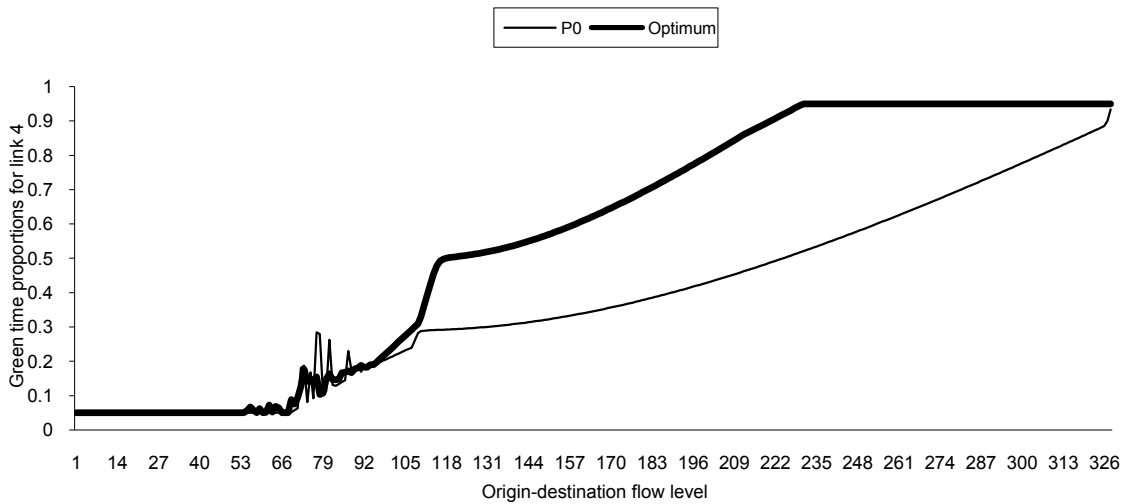


Figure 7. Green-time proportions awarded to link 4, versus origin-destination flow level, for both control methods.

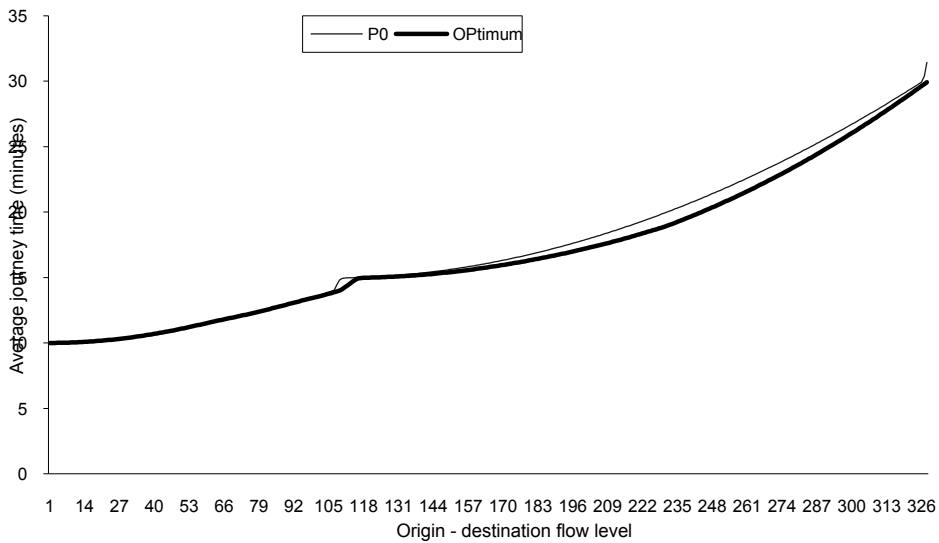


Figure 8. Average journey time (minutes), versus origin-destination flow level, for both control methods.

The significance of these results is that the total cost at equilibrium using P_0 is throughout the demand range only just slightly greater than the total cost at the hard to find optimum signal timings. The P_0 timings are close to the optimum timings when the demand approaches the network capacity.

6.1 Application to the design of intelligent interventions

The results above suggest a method of designing intelligent changes in fixed time signal timings to suit different scenarios; possibly brought about by an incident. Suppose that the total origin-destination flow through the network in figure 6 changes from level 100 to level 300. Then if the P_0 results here were available and being followed, they would suggest recommending that the green-time proportion awarded to link 4 is changed to about 0.85 (possibly

from about 0.25). Provided the modelling has already been done this change might be recommended quickly. Designing fixed time signal settings in this way has been suggested in Smith (2010).

7. Conclusions

7.1 Modelling Conclusions

The paper has considered a simple day to day dynamic re-routeing process suggested in Smith (1984a); this uses route-swaps. He et al (2010) have shown that this route-swapping process is unrealistic and in this paper we show that this unrealism may be removed by using splitting rates rather than route flows in the swapping process. A rigorous proof of a simple convergence or stability result in this splitting rate framework is given in appendix D.

The paper extends the splitting rate re-routeing framework so as to include the dynamic re-assignment of green time (using the P_0 policy) in addition to traffic flows. An example is given which compares signal timings which arise using this method and signal timings which arise when bilevel optimisation is employed. Also, to set the assignment / control scene, an initial example has been given where the equisaturation policy is utilised with a natural re-routeing algorithm; very unpredictable behaviour, including the pitchfork bifurcation discontinuity and hysteresis, arise.

Further modelling work may be helpful, in several directions. Three of these are as follows.

1. To investigate the efficiency of the splitting rate method as an equilibrium-calculating method.
2. To consider whether the splitting rate approach here can be adapted to suit more general networks, especially networks with loops. (In this paper we have made assumptions that prevent looping.)
3. To consider whether it is necessary for a convergence proof that there is flow out of every non-destination node. (The particular convergence proof given in the appendix here certainly does need this assumption.)

7.2 Conclusions concerning real-life traffic control

The paper highlights two concerns related to the real life responsive control of urban road networks. On one hand the initial example shows that the interaction between responsive control and routeing, when the familiar equisaturation policy is utilised, may lead to undesirable discontinuities and uncertainties. On the other hand the paper shows that these undesirable outcomes do not (under certain conditions) arise when the P_0 policy is utilised responsively. Further work may be desirable in certain “real-life” directions. Two of these are as follows: (i) to establish whether instabilities like those in section 1 here actually occur in real life networks controlled by responsive control systems; and (ii) to estimate the likely efficiency of signal settings using the P_0 policy.

Acknowledgments

1. The authors of this paper have received support via the FREEFLOW project; this support is gratefully acknowledged. FREEFLOW is aimed at the development of systems which transform road transport data into intelligence; partly to help network operators and managers make better (and more timely) signal control interventions. The partners in FREEFLOW are: Transport for London, City of York Council, Kent County Council, ACIS, Kizoom, Mindsheet, QinetiQ, Trakm8, Imperial College London, Loughborough University and the University of York. FREEFLOW is funded by the Technology Strategy Board, the Department for Transport, EPSRC and the partners.

2. Section 2 of this paper was joint work with David Slade while he was a student on the MSc course “Mathematics with Modern Applications” at the University of York.

3. We acknowledge helpful comments by Ian Routledge on real-life signal control systems.

4. The paper has benefitted substantially from the comments of anonymous referees.

References

- Aashtiani, H., Magnanti, T., 1983. Equilibrium on a congested transport network. *SIAM Journal of Algebraic and Discrete Methods*, 2, 213-216.
- Allsop, R. E., 1974. Some possibilities for using Traffic Control to Influence Trip Distribution and Route Choice. *Proceedings of the Sixth International Symposium on Transportation and Traffic Theory*, Sydney (Editor: Buckley, D. J.), Elsevier, 345-374.
- Bar-Gera, H., 2002. Origin-based algorithm for the traffic assignment problem. *Transportation Science* 36 (4), 398 – 417.
- Bar-Gera, H., Boyce, D., 2003. Origin-based algorithms for combined travel forecasting models. *Transportation Research Part B*, 37, 405-422.
- Bar-Gera, H., Boyce, D., 2006. Solving a non-convex combined travel forecasting model by the method of successive averages with constant step sizes. *Transportation Research Part B* 40, 351-367.
- Bar-Gera, H., 2009. Traffic Assignment by Paired Alternative Segments. Preprint. Ben Gurion University.
- Beckmann, M., McGuire, C.B., Winsten, C. B., 1956. *Studies in the Economics of Transportation*. Yale University Press, New Haven, CT.
- Bertsekas, D., Gallager, R., 1992. *Data networks* (2nd edition). Prentice Hall.
- Bie, J., Lo, H. K., 2010. Stability and attraction domains of traffic equilibria in a day-to-day dynamical system formulation. *Transportation Research Part B* 44, 90 – 107.
- Cantarella, G. E., 1997. A general fixed-point approach to multimodal multi-user equilibrium assignment with elastic demand. *Transportation Science* 31 (2), 107-128.
- Cantarella, G., Cascetta, E., 1995. Dynamic processes and equilibrium in transportation networks: Towards a unifying theory. *Transportation Science* 29 (4), 305 – 329.
- Charnes, A., Cooper, W. W., 1958. Multicopy traffic network models. *Proceedings of the Symposium on the Theory of Traffic Flow*, General Motors Research Laboratories (Editor: R. Herman), 1958.
- Clegg, J., Smith, M. J., Xiang, Y., Yarrow, R., 2001. Bilevel programming applied to optimising urban transportation. *Transportation Research Part B*, 35, 41-70.
- Dafermos, S., 1980. Traffic equilibrium and variational inequalities. *Transportation Science* 14, 42 – 54.
- Dial, R., 1971. A probabilistic multipath assignment model that obviates path enumeration. *Transportation Research* 5, 83-111.
- Dial, R., 2006. A path-based user-equilibrium assignment algorithm that obviates path storage and enumeration. *Transportation Research Part B* 40, 917 – 936.
- Dickson, T. J., 1981. A note on traffic assignment and signal timings in a signal-controlled road network. *Transportation Research Part B*, 267 – 271.
- Evans, S. P., 1976. Derivation and analysis of some models for combining trip distribution and assignment. *Transportation Research* 10, 37-57.
- Gartner, N. H., 1983. OPAC: A demand responsive strategy for traffic signal control. *Transportation Research Record* 906, 75 - 81.
- Gentile, G., 2009. On the accurate convergence of deterministic assignment when comparing scenarios for large networks; investigating the LUCE algorithm. Paper presented at the SIDT 2009 International Conference (Milan, June).
- Gentile, G., Noekel, K., 2009. Linear user cost equilibrium: the new algorithm for traffic assignment in VISUM. Paper presented at the European Transport Forum (Brussels, September).
- He, X., Guo, X., Liu, H., 2010. A link-based day to day traffic assignment model. *Transportation Research Part B*, 44, 597 – 608.
- Heydecker, B. G., Chen, C., Wong, C.K., 2007. Adaptive dynamic control for road traffic signal. *IEEE International Conference on Networking, Sensing and Control*, London.
- Heydecker, B. G., 2004. Objectives, stimulus and feedback in signal control of road traffic. *ITS Journal* 8(2), 63-76
- Hunt, P. B., Robertson, D. I., Bretherton, R. D., 1982. The SCOOT on-line traffic optimisation technique. *Traffic Engineering and Control*, 190 – 192.
- Kupiszewska, D., Van Vliet, D., 1999. 101 uses for path-based assignment. *Proceedings of the European Transport Conference*, PTRC Education and Research Services Ltd. Seminar F, *Transportation Planning Methods*, 121-132.
- Larsson, L.J., Patriksson, M., 1992. Simplicial decomposition with disaggregated representation for the traffic assignment problem. *Transportation Science* 26 (1), 4-17.

- LINSIG (2010). <http://en.wikipedia.org/wiki/LINSIG> (downloaded on 14.08.10).
- Lv, Q., Salis, A., Skrobanski, G., Smith, M.J., Springham, J., Woods, A., Gordon, A., 2007. Two-direction methods for variable demand traffic assignment. *Mathematics in Transport (Selected Proceedings of the 4th IMA International Conference on Mathematics in Transport)*. (Editor: B. Heydecker), Elsevier, 1-96.
- Lyapunov, A. M., 1907. Probleme general de la stabilitate de mouvement, Ann. Fac. Sci. Toulouse 9, 203 – 474. Reprinted in Ann. Math. Stud. No 17 (Princeton University Press), 1949.
- Payne, H. J., Thompson, W. A., 1975. Traffic assignment on transportation networks with capacity constraints and queueing. Paper presented at the 47th National ORSA/TIMS North American Meeting.
- Maher, M. J., 1992. SAM – a stochastic assignment model. *Mathematics in Transport Planning and Control (Based on the the Proceedings of the Second IMA International Conference on Mathematics in Transport Planning and Control)* (Editor J D Griffiths), Oxford University Press, 121 - 143.
- Maher, M. J., Hughes, P. C., 1997. A probit-based stochastic user equilibrium assignment model. *Transportation Research Part B*, 31, 341 – 355.
- Maher, M., 1998. Algorithms for logit-based stochastic user equilibrium assignment. *Transportation Research Part B* 32, 539 – 549.
- Mounce, R., 2009. Existence of equilibrium in a continuous dynamic queueing model for traffic networks with responsive signal control. Proceedings of the Eighteenth International Symposium on Transportation and Traffic Theory, Hong Kong Polytechnic University, Hong Kong. (Editors: Lam, W. H. K., Wong, S. C., Lo, H. K.), Springer, 327-344.
- Nie, Y., 2010. A class of bush-based algorithms for the traffic assignment problem. *Transportation Research Part B* 44(1), 73 – 89.
- Patriksson, M., 1994. *The traffic assignment problem – Models and Methods*. VSP, Utrecht, Netherlands.
- Robertson, D. I., 1969. *TRANSYT: a traffic network study tool*. RRL Lab. report LR253, Road Research Laboratory, Crowthorne, UK.
- Sheffi, Y., 1985. *Urban transportation Networks: Equilibrium Analysis with Mathematical Programming Methods*, Prentice Hall (New Jersey) and available on-line.
- Smith, M. J., 1979. A local traffic control policy which automatically maximises the overall travel capacity of an urban road network. Proceedings of the International Symposium on Traffic Control Systems, University of California, Berkeley, Part 2A, 11-32; and in *Traffic Engineering and Control*, 21, 1980, 298-302.
- Smith, M. J., 1984a. The stability of a dynamic model of traffic assignment - an application of a method of Lyapunov. *Transportation Science* 18, 245 - 252.
- Smith, M. J., 1984b. A descent algorithm for solving a variety of monotone equilibrium problems. Proceedings of the Ninth International Symposium on Transportation and Traffic Theory, Delft University of Technology, Netherlands (Editors: Volmuller, J., Hamerslag, R.), VNU Science Press, 273-297.
- Smith, M. J., 1984c. A descent method for solving monotone variational inequalities and monotone complementarity problems. *Journal of Optimization Theory and Applications*, 44, 485-496.
- Smith, M. J., 1987. Traffic control and traffic assignment in a signal-controlled network with queueing. Proceedings of the Tenth International Symposium on Transportation and Traffic Theory, MIT, Cambridge, Massachusetts (Editors: Gartner, N. H., Wilson, N. H. M.), Elsevier, 61-68.
- Smith, M. J., van Vuren, T., 1993. Traffic equilibrium with responsive traffic control. *Transportation Science*, 27, 118-132.
- Smith, M. J., 2006. Bilevel optimisation of prices and signals in transportation models. *Mathematical and Computational Models for Congestion Charging* (Editors: Lawphongpanich, S., Hearn, D. W., Smith, M. J.), 159 – 199.
- Smith, M. J., 2009. A Two-direction Method of Solving Variable Demand Equilibrium Models with and without Signal Control. Proceedings of the Eighteenth International Symposium on Transportation and Traffic Theory, Hong Kong Polytechnic University, Hong Kong. (Editors: Lam, W. H. K., Wong, S. C., Lo, H. K.), Springer, 365 - 386.
- Smith, M. J., 2010. Intelligent Network Control: using an Assignment-Control Model to Design Fixed Time Signal Timings. *Transport Planning – Advances in Dynamic Traffic Assignment* (Editors: Tampere, C., Viti, F., Immers, L.), Edward Elgar, 57 – 72.
- Van Vuren, T., Van Vliet, D., 1992. *Route Choice and Signal Control*. Avebury.
- Wardrop, J. G., 1952. Some Theoretical Aspects of Road Traffic Research, Proceedings, Institution of Civil Engineers II (1), 235-278.

- Watling, D., 1996. Asymmetric problems and stochastic process models of traffic assignment. *Transportation Research Part B* 30(5), 339 – 357.
- Watling, D., 1999. Stability of the stochastic assignment problem, a dynamical systems approach. *Transportation Research Part B* 33, 281 – 312.
- Watling, D., Hazelton, M., 2003. The dynamics and equilibria of day-to-day assignment models. *Networks and Spatial Economics* 3, 349 – 370.
- Webster, F. V., 1958. *Traffic Signal Settings*, Department of Transport, Road Research Technical Paper No. 39, HMSO, London.
- Wood, K., 1993. *Urban Traffic Control, Systems Review*. Project Report 41, Transport Research Laboratory, Crowthorne, United Kingdom.
- Yang, H., Yagar, S., 1995. Traffic assignment and signal control in saturated road networks. *Transportation Research Part A* 29, 125 - 139.
- Yang, H., 1996. Equilibrium network traffic signal setting under conditions of queueing and congestion. *Applications of Advanced Technologies in Transportation Engineering*. (Editors: Stephanedes, Y. J., Filippi, F.), American Society of Civil Engineers, 578 - 582.

APPENDICES

In these appendices we give a short context within which the contents of this paper lie (A, B and C) and a rigorous proof of convergence of the splitting rate algorithm (under certain conditions) to a set of approximate equilibria in a finite number of steps (D).

Appendix A. Signal Control Context

Webster (1958) considered ways of determining signal timings for a single isolated intersection. As a result of his theoretical and simulation studies he suggested that the equisaturation policy would be a practical way of approximately minimising the total rate of delay to vehicles passing through the intersection. If the intersection has just two approaches then this equisaturation policy aims to choose signal green times so that the saturation ratios on the two approaches are equal. [If there are more than two approaches and many signal stages the policy considers the most saturated approach within each stage (the “representative” approach for that stage) and aims to choose green-times which equalize the degrees of saturation of these “representative” approaches.] LINSIG (2010) is now often used to design or determine signal timings at a single signal-controlled intersection, and also over small networks. The assumption here is that a single set of timings, once designed, will be applied in an unchanging manner. Thus they are called “fixed-time” signal settings. Different fixed time settings may be utilized at different times of day.

There are now several models in use for designing or optimizing signal timings over a whole network. The most well-known is TRANSYT (TRAffic Network StudY Tool; Robertson 1969); this may be used to design fixed timings, where the timings do not respond rapidly to the prevailing traffic flows. Unlike Webster’s method signal timings designed using TRANSYT do allow for adjacent junctions or (for example) for a sequence of junctions on one main route. In TRANSYT the whole network and the bottlenecks within it have an impact on the signal design process and on the timings suggested.

Adaptive or responsive systems seek to adapt signals timings in near to real time in response to changing traffic flows and include: SCOOT (Split, Cycle and Offset Optimisation Technique; Hunt, 1982; this started as a responsive version of TRANSYT and follows equisaturation rules similar to Webster’s for deciding how green time is split among stages); SCATS (Sydney Co-ordinated Adaptive Traffic System; UTOPIA (developed by FIAT, Mizar and others; see www.miz.it) and OPAC (see Gartner (1983). See Wood (1993) for a helpful discussion of the various systems. Heydecker et al (2007) and Heydecker (2004) outline motivations and new approaches to adaptive signal control. Key practical issues for real-life responsive traffic control systems include:

- (a) UTC plan structure;
- (b) use of demand dependent stages, including ending a stage early if there is a gap in the traffic; and
- (c) maintaining good detection in suitable locations despite bad weather, wear and tear and road maintenance.

The control variables considered in both fixed time and responsive systems include not only how the total green time is split between stages at each junction (the *splits*) but also *offsets* which determine how display changes at different signals are related and *cycle times* which determine for each signal the time which must elapse before the signal display repeats. Offsets are often adjusted to aid progression through a sequence of closely spaced signals – providing a *green wave*; although a *red wave* is also possible to deter excessive use of a sensitive route. Cycle time is normally increased when congestion is high so as to reduce to a minimum the time “lost” due to the clearance times between stages; these clearance times depend on geometrical factors and are fixed (determined by safety considerations); therefore a longer cycle time means that lost times make up a smaller proportion of each cycle. In the UK, for most networks the longest cycle time is usually 120 seconds.

A central divide in real-life traffic control systems is between systems which are essentially fixed time and systems which are essentially responsive. A particular fixed time system tuned to benefit just one pattern of traffic may well be less appropriate for a changed traffic pattern; changed patterns may arise slowly over time or suddenly due to an incident. In fixed time systems tailored plans for specific situations are ideally required. On the other hand a fixed time control system requires rather less on-line data communication, data analysis and maintenance than responsive or adaptive systems.

Appendix B. Routeing Context

The day to day dynamics of traffic re-routeing has been considered by Bie and Lo (2010), Cantarella and Cascetta (1995), Smith (1984a), Watling (1996, 1999) and Watling and Hazelton (2003) and others. These papers consider deterministic systems and stochastic networks and do not involve splitting rates.

Traffic assignment has been considered in a vast number of papers and books. These papers consider the design of iterative methods which ensure that models converge to equilibrium; without seeking to design the iterations so as to represent a realistic within-day or day to day dynamical system. The following, selecting a minute proportion of the literature, merit consideration. Bar-Gera and Boyce (2003), Bar-Gera and Boyce (2006), Cantarella (1997), Charnes and Cooper (1958), Dafermos (1980), Dial (2006), Dial (1971), Evans (1976), Larsson and Patriksson (1992), Lv et al (2007), Maher (1992), Maher and Hughes (1997), Maher (1998), Patriksson (1994), Payne and Thompson (1975), Sheffi (1985), Smith (1984b, c) and Smith (2009). Some of these papers are also concerned with models where demand is variable.

Appendix C. Simultaneous Routeing and Signal Optimisation

The effects of the changing responsive signal timings on route-choices and other decisions by users are often ignored by signal control designers. It was first pointed out by Allsop (1974) that signal timings should ideally take reasonable account of the reactions of travellers; this is partly to try to optimize signals subject to an equilibrium constraint (at which all travellers are happy with their route-choices) and partly to at least obtain a consistent (green-time, route-flow) pair (G^* , X^*), say. If for example the equisaturation policy is to be employed then it is reasonable to seek green-times and flows where (1) green-times satisfy the equisaturation policy at each junction and (2) for each origin-destination pair no traveller has a less costly alternative route. The latter is Wardrop’s equilibrium condition. Dickson (1981) first showed that using delay-minimising signal settings does not minimize delay at a Wardrop equilibrium.

Typically existing control systems seek to be conservative so as not to generate disturbances by over-reaction; however dynamical routeing / control issues have been little studied from either a theoretical or practical viewpoint and so merit much more consideration.

Much of the existing theoretical work on re-routeing / control interactions has focused on one particular dynamical system which can only occur within models: this is the standard method of trying to achieve a consistent (green-time, flow) pair (G^* , X^*). The method iterates between the signal setting model (determining exactly equisaturating green-times G for fixed flows X) and the traffic assignment model (determining exactly equilibrium route-flows X for fixed signal settings G). This is called Iterative Optimisation Assignment (IOA). Convergence of the IOA dynamical system has only been proved for a few control policies; see Smith and Van Vuren (1993) for some rather recondite stability results involving responsive control. Smith (1987) considered the assignment / control problem with explicit queues. Combining signal control and route choice has been considered by some others: see Yang (1996) and Yang and Yagar (1995), for example. In all the above work the setting is static within a

day. Recently Mounce (2009) has considered the problem of existence of equilibrium in a continuous dynamic queueing model for traffic networks with responsive signal control, in a dynamic within day setting.

In this paper we give an example to show that routing / control dynamics may give rise to instability and unpredictability; and we also outline a model within which routing / control dynamical systems may be studied. This is a very natural and simple extension of the traffic assignment splitting rate model.

The paper also gives a more positive example showing how the methods outlined in the paper may be used to generate signal control interventions to meet different circumstances.

Appendix D. Proof of convergence to a set of approximate equilibria in a finite number of steps under certain conditions

This proof continues the arguments in section 4 in the main body of the paper. We retain the single-destination loop-free network with fixed demands emanating from every non-destination node. Also as before D is the set of link flow vectors consistent with the fixed demands.

If each link cost is a continuous function of only the flow along the link, the set E of equilibria is given below (where Z is given in (30)).

$$E = \{ \mathbf{x} \in D; - \text{grad } Z(\mathbf{x}) = - \mathbf{c}(\mathbf{x}) \text{ is normal at } \mathbf{x} \text{ to } D \}. \tag{D1}$$

For $\varepsilon > 0$, let

$$E_\varepsilon = \{ \mathbf{x} \in D; \text{dist}(\mathbf{x}, E) < \varepsilon \} \text{ and } F_\varepsilon = \{ \mathbf{x} \in D; \text{dist}(\mathbf{x}, E) \geq \varepsilon \} \tag{D2}$$

where $\text{dist}(\mathbf{x}, E)$ is the Euclidean distance between \mathbf{x} and the equilibrium set E in (D1).

Then we have the following result; which utilises results in section 4.4 and the set E_ε just defined above in (D2).

Theorem 1. (Convergence to the set E_ε of approximate equilibria.)

Suppose given

- (a) a loop-free single destination network;
- (b) for each non-destination node, a fixed or rigid positive demand from each non-destination node to the single destination node; and
- (c) for each link i a continuous and positive link cost function $c_i(\cdot)$, defined for all $x_i \geq 0$.

Let D be the set of (non-negative) link flow vectors \mathbf{x} consistent with the given fixed positive demands, let the direction $\Delta\mathbf{x}$ be generated by $\delta\mathbf{x}$ following (23), (24), and (25) or (28), and let k be sufficiently small to ensure that

$$\mathbf{x} \geq \mathbf{0} \Rightarrow \mathbf{x} + \Delta\mathbf{x} \geq \mathbf{0}$$

for all link flow vectors $\mathbf{x} \in D$.

Let $\varepsilon > 0$, $0 < h < 1$ and let

$$\mathbf{x}^1, \mathbf{x}^2 = \mathbf{x}^1 + h\Delta\mathbf{x}^1, \mathbf{x}^3 = \mathbf{x}^2 + h\Delta\mathbf{x}^2, \dots, \mathbf{x}^{n+1} = \mathbf{x}^n + h\Delta\mathbf{x}^n, \dots \tag{D3}$$

be an infinite sequence in D .

Then if h is sufficiently small the infinite sequence (D3) eventually enters E_ε .

Proof. (In this proof dist stands for Euclidean distance.)

Let $\mathbf{x} \notin E_\varepsilon$ or $\mathbf{x} \in F_\varepsilon$. Then by section 4.4, $\Delta\mathbf{x} \neq \mathbf{0}$ and using (33) and (34),

$$[\Delta\mathbf{x}]^T \text{grad } Z(\mathbf{x}) < 0.$$

Now $[\Delta\mathbf{x}]^T \text{grad } Z(\mathbf{x})$ is a continuous function of $\mathbf{x} \in F_\varepsilon$ and F_ε is compact (closed and bounded). Therefore there is a constant $c > 0$ such that

$$[\Delta\mathbf{x}]^T \text{grad } Z(\mathbf{x}) < -c$$

for all $\mathbf{x} \in F_\varepsilon$.

Also $[\Delta\mathbf{x}]^T \text{grad } Z(\mathbf{y})$ is a continuous function of $(\mathbf{x}, \mathbf{y}) \in D \times D$ and $D \times D$ is compact. (It is the product of two compact sets.) Hence $[\Delta\mathbf{x}]^T \text{grad } Z(\mathbf{y})$ is a uniformly continuous function of (\mathbf{x}, \mathbf{y}) on $D \times D$ and there must be $a > 0$ such that

$$[\Delta\mathbf{x}]^T \text{grad } Z(\mathbf{y}) - [\Delta\mathbf{x}]^T \text{grad } Z(\mathbf{x}) < c/2 \text{ if } \text{dist}((\mathbf{x}, \mathbf{x}), (\mathbf{x}, \mathbf{y})) = \text{dist}(\mathbf{x}, \mathbf{y}) < a.$$

Further, since $\Delta\mathbf{x}$ varies continuously with \mathbf{x} , there is $h > 0$ such that

$$\text{dist}(\mathbf{x}, \mathbf{x} + t\Delta\mathbf{x}) = \text{dist}(\mathbf{0}, t\Delta\mathbf{x}) < a \text{ if } \mathbf{x} \in D \text{ and } 0 \leq t \leq h.$$

Hence, if $\mathbf{x} \in D$ and $0 \leq t \leq h$,

$$[\Delta\mathbf{x}]^T \text{grad } Z(\mathbf{x} + t\Delta\mathbf{x}) - [\Delta\mathbf{x}]^T \text{grad } Z(\mathbf{x}) < c/2.$$

Suppose now that $\mathbf{x} \in F_\varepsilon$ and $0 \leq t \leq h$. Then it follows from above that:

$$\begin{aligned} [\Delta\mathbf{x}]^T \text{grad } Z(\mathbf{x} + t\Delta\mathbf{x}) &= [\Delta\mathbf{x}]^T \text{grad } Z(\mathbf{x}) + \{[\Delta\mathbf{x}]^T \text{grad } Z(\mathbf{x} + t\Delta\mathbf{x}) - [\Delta\mathbf{x}]^T \text{grad } Z(\mathbf{x})\} \\ &< \quad \quad \quad -c \quad \quad \quad + \quad \quad \quad c/2 \\ &= \quad \quad \quad -c/2. \end{aligned}$$

It now follows that

$$Z(\mathbf{x} + h\Delta\mathbf{x}) - Z(\mathbf{x}) < -ch/2$$

for all $\mathbf{x} \in F_\varepsilon$.

Hence, if $\mathbf{x}^1, \mathbf{x}^2, \mathbf{x}^3, \dots, \mathbf{x}^{n-1}, \mathbf{x}^n$ (the first n terms of the sequence (D3)) belong to F_ε , then

$$\begin{aligned} Z(\mathbf{x}^n) &= Z(\mathbf{x}^1) + [Z(\mathbf{x}^2) - Z(\mathbf{x}^1)] + \dots + [Z(\mathbf{x}^{n-1}) - Z(\mathbf{x}^{n-2})] + [Z(\mathbf{x}^n) - Z(\mathbf{x}^{n-1})] \\ &< Z(\mathbf{x}^1) - ch/2 - \dots - ch/2 - ch/2 \\ &= Z(\mathbf{x}^1) - (n-1)ch/2. \end{aligned}$$

This is negative if n is large, which is not possible as Z must by definition be non-negative (since the cost functions are all positive). Therefore the sequence (D3) must leave F_ε for some n or must enter F_ε for some n .

Remark 1. The above result may be easily modified to ensure that the sequence enters and remains forever in E_ε , for some n .

Remark 2. Convergence of the MSA and similar algorithms in this splitting rate setting follows very quickly from the arguments used here, by using instead of (D3) the sequence

$$\mathbf{x}^1, \mathbf{x}^2 = \mathbf{x}^1 + h_1\Delta\mathbf{x}^1, \mathbf{x}^3 = \mathbf{x}^2 + h_2\Delta\mathbf{x}^2, \dots, \mathbf{x}^{n+1} = \mathbf{x}^n + h_n\Delta\mathbf{x}^n, \dots \tag{D4}$$

where $h_n \rightarrow 0$ as n tends to infinity, and $\sum_i h_i$ diverges. With this change, but otherwise under the same conditions

as above, for any ε the sequence (D4) enters E_ε .