Article:

Liu, R and Smith, MJ (2015) Route choice and traffic signal control: a study of the stability and instability of a new dynamical model of route choice and traffic signal control. Transportation Research Part B: Methodological, 77. 123 - 145. ISSN 0191-2615

https://doi.org/10.1016/j.trb.2015.03.012

Please cite this article in press as: Liu, R. and Smith, M. J. (2015) Route choice and traffic signal control: A study of the stability and instability of a new dynamical model of route choice and traffic signal control. **Transportation Research Part B**. Vol 77, pp123-145. DOI: http://dx.doi.org/10.1016/j.trb.2015.03.012

Route choice and traffic signal control: A study of the stability and instability of a new dynamical model of route choice and traffic signal control

Ronghui Liu[a], Mike Smith[b]

[a] Institute for Transport Studies, University of Leeds, Leeds LS2 9JT, UK
Email: r.liu@its.leeds.ac.uk
[b] Department of Mathematics, University of York, York YO10 5DD, UK
Email: 152mikesmith@gmail.com

ABSTRACT

This paper presents a novel idealised dynamical model of day to day traffic re-routeing (as traffic seeks cheaper routes) and proves a stability result for this dynamical model. (The dynamical model is based on swapping flow between paired alternative segments (these were introduced by Bar-Gera (2010)) rather than between routes.) It is shown that under certain conditions the dynamical system enters a given connected set of approximate equilibria in a finite number of days or steps. This proof allows for saturation flows which act as potentially active flow constraints. The dynamical system involving paired alternative segment swaps is then combined with a novel green-time-swapping rule; this rule swaps green-time toward more pressurised signal stages. It is shown that if (i) the delay formulae have a simple form and (ii) the "pressure" formula fits the special control policy P_0 (see Smith, 1979a, b), then the combined flow-swapping / green-time-swapping dynamical model also enters a given connected set of approximate consistent equilibria in a finite number of steps. Computational results confirm, in a simple network, the positive P_0 result and also show, on the other hand, that such good behaviour may not arise if the equi-saturation control policy is utilized. The dynamical models described here do not represent blocking back effects.

Key words: Dynamics, Convergence, Stability, Routeing, Signal control, Day to day

1. Introduction

1.1 **A brief route choice and traffic control modelling context**

Dynamic transport models involving both travellers' choices (including drivers' repeated route choices) and traffic signal controls are needed. Such models may be used

(i) to help predict (for a given responsive control strategy) how traffic flows and controls are likely to evolve over time and so to help assess different given control strategies (against specified congestion, delay, pollution, accessibility or other criteria); and

(ii) to help design new control strategies for reducing congestion, delay, pollution, inaccessibility (or other criteria) in cities, taking reasonable account of the future evolution of traffic flows as these respond to the control strategies.

Allsop (1974), Gartner (1976), Smith (1979a, c), Bentley and Lambe (1980) and Dickson (1981) were among the first to point to the need to combine models of route choice and traffic signal control; in part so that optimal controls taking account of routeing reactions might be found. The study of traffic control and route choice has been pursued by Meneguzzer (1996, 1997), Maher et al (2001), Wong et al (2001), and many others. Taale and van Zuylen (2001) provide an overview.

Cantarella et al (1991) and Cantarella (2010) focus on seeking optimal controls which take account of route choices. They address stability issues involving both routeing and control. In these papers, a bi-level optimisation method is used as the signal setting method. The route choice model used finds for each OD pair a cheapest route and then swaps route flow toward the cheapest route.

In this paper we consider a joint, two-commodity (route flow, green time) dynamical system; in which route-flows switch toward cheaper routes and signal green-times switch to more pressured stages. Both route-flow and green-time swaps follow a development of the 'proportional adjustment process' dynamical system in Smith (1984a).

In combined traffic signal control and route-choice models we consider not only costs of routes (which will causes route flows to change) but also "pressures" on signal stages (which will cause stage green-times to change). This formulation was perhaps first introduced in Smith et al (1987) and Smith (1987). In the models here both route costs and stage pressures will be functions of flows and green-times. These given functions determine (flow, green-time) pairs which satisfy Wardrop's equilibrium condition and a specific control policy as follows:

A (route-flow, green-time) pair satisfies the Wardrop equilibrium condition if:

more costly routes carry no flow. (1.1)

A (route-flow, green-time) pair satisfies the signal control policy if:

less pressurised stages receive no green-time. (1.2)

Condition (1.1) holds if and only if Wardrop's equilibrium condition is exactly satisfied. If (1.1) does not hold exactly then it is proposed initially in this paper that for each pair of routes joining each OD pair, route flow swaps from the more costly route to the less costly route at a rate which is proportional to:

(the difference in the route costs) × (flow along the route with the greater cost).

Similarly, condition (1.2) holds if and only if the control policy is exactly satisfied. If (1.2) does not hold exactly then, in this paper, for each pair of stages at each junction the stage green-time swaps from the less pressurised stage to the more pressurised stage at a rate which is proportional to:

(the difference in the stage pressures) × (green-time given to the stage with the smaller pressure).

The stability of this combined routeing and signal-control dynamical system is considered in this paper. Smith and Mounce (2011) have considered a restricted form of this dynamical model within a very different context: that of splitting rates.

A wide-ranging route choice and signal control modelling context is given in Appendix A.

1.2 **Overview and contributions of this paper**

This paper focuses on certain mathematical models of route choice dynamics and combined route-choice and traffic signal control dynamics and considers the stability of these dynamical models.

The routeing plus signal control dynamical system represents car drivers seeking better routes and signal timing changes in response to changing traffic flows. The combined (routeing, signal control) dynamical model is idealized. The signal control model may be regarded as a model of a system periodically updated either by an operator or by an automatic system. The dynamical routeing model is designed to approximately represent, albeit in a simplified or idealized form, how routeing decisions are actually made day after day.

The first contribution of this paper is to introduce a new route choice dynamical system; this is a restricted version of the proportional-switch adjustment process (or PAP) suggested in Smith (1984a) and discussed by He et al (2010). He et. al. (2010) show that the PAP route-swapping model is not always realistic. In this paper we put forward a restricted proportional adjustment process (RPAP) to take account of the comments by He et al (2010) while maintaining the essential (proportional) characteristics of the PAP. Paired alternative segments (introduced by Bar-Gera (2010)) form a central element of a RPAP.

Other route swap algorithms have been considered by Cascetta (1989), Smith and Wisten (1995), Bellei et al. (2005), Huang and Lam (2002), Peeta and Yang (2003), Nie and Zhang (2005), Nie (2010), Mounce (2006, 2009), Mounce and Carey (2011) and Mounce and Smith (2007). None utilise paired alternative segments.

Secondly, we show that the above route-swapping dynamical system satisfies a stability property similar to that already proved for the more artificial PAP dynamical re-routeing system described in Smith (1984a). To be precise it is shown that, under natural conditions, a trajectory of the route-flow dynamical system enters a set of approximate equilibria in a finite number of "days". This routeing stability is guaranteed using RPAP and a discrete dynamical system with fixed step lengths instead of PAP and a smooth solution to a differential equation, and moves the initial differential PAP theory in Smith (1984a) towards both computer implementation and reality.

A third contribution of the paper is to add to the proposed new RPAP re-routeing model a corresponding dynamical model of signal control. To do this most simply we introduce signal red-times as the control variable (in place of green-times); and think of the red-time allocated to a signal-controlled link (the proportion of time a link is "red") as an extra 'flow' through that signal controlled link.

We show that Webster's equi-saturation policy (Webster, 1958) and the P_0 control policy introduced in Smith (1979a, b), may then be readily included within the dynamical RPAP route-flow swapping process; by simultaneously swapping red times at each junction and route-flows joining each OD pair. While route flows are swapped between certain pairs of routes according to certain costs and cost differences, red times are swapped according to certain "pressures" and pressure differences; different definitions of these pressures then give rise to dynamical versions of different control policies.

Finally, a fairly general stability result is proved for the central [route-flow, red-time] swapping model; this is shown to hold when routeing and signal controls vary simultaneously, provided that the responsive signal control policy P_0 is utilized and provided that the delay formula involving both flow and red times has a certain form; this form ensures that the delay felt at a link exit is a non-increasing function of the spare capacity at that link exit.

To show that the above stability guarantee is a property of the control policies utilised and the dynamical assumptions made, and is not a result which holds generally, a simple example is given. This example shows that if the equi-saturation policy (rather than P_0) is combined with re-routing then convergence to a single connected equilibrium set no longer holds: in this simple example (with equi-saturation) there is a pitchfork bifurcation and unpredictable behaviour arises, including hysteresis.

The paper is organised as follows. In Section 2, the restricted proportional-switch adjustment process (RPAP) for dynamical route-choice modelling is introduced and stability results are obtained. Section 3 introduces the extension of RPAP to combined dynamical route-flow swaps and red-time swaps, and presents stability results for the combined system. Example numerical results displaying instability with the equi-saturation policy are presented in Section 4. Finally, Section 5 concludes the paper.

2. The Restricted Proportional-Switch Adjustment Process (RPAP) for Route Swaps

Smith (1984a) proposed a simple day-to-day re-routeing process (called a "proportional-switch adjustment process", or PAP, by He et. al., 2010). Three purposes were identified: (i) to allow the stability or otherwise of a given traffic equilibrium to be studied (unstable equilibria are unlikely to persist); (ii) to help determine, when there are several equilibria, which equilibrium attracts trajectories starting from a given point; and (iii) to allow the possibility of modelling moving a traffic equilibrium to another "better" equilibrium by using a (perhaps temporary) signal-control intervention.

He et al (2010) highlight a behavioural deficiency of this dynamical route-swap model; this deficiency arises in part because of route overlaps. In this paper we show how this behavioural deficiency of PAP may be removed by considering route swaps which are more restricted than those in PAP, but keeping the proportionality of PAP. This new "restricted" route-swapping version of PAP seems likely to be the simplest way of "correcting" PAP in light of the observation of He et al (2010). It is clear that the anomaly identified by He et al (2010) cannot arise with the dynamical re-routeing model presented in this paper.

We present below the original PAP as proposed by Smith (1984a) for a simple network in section 2.1, before introducing the RPAP in Sections 2.2 and 2.3. Stability results for RPAP are presented in Sections 2.4 and 2.5.

2.1 The proportional adjustment process (PAP) in a simple network

Suppose that travellers traverse the small network illustrated in figure 1, and that these same travellers traverse the network repeatedly. In this paper to be definite we think of this repetition as “day after day” (although we might also think more generally in terms of “epoch to epoch”). 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?

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 on day t swapping to route 2 on day t + 1; this would often oscillate from day to day and is, under normal circumstances, 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.

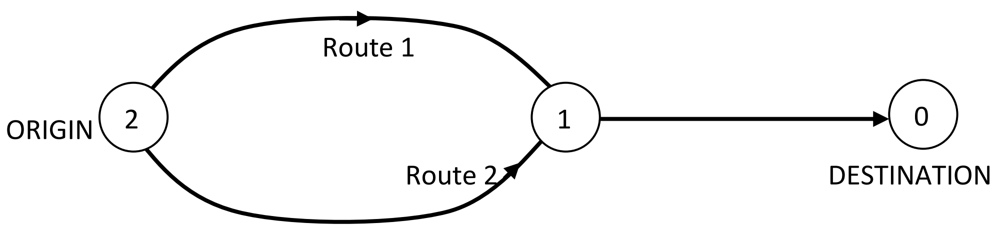

Figure1. A two route network.

It is not clear how many travellers will, in reality, swap for a given actual or perceived cost difference. A natural and simple assumption is that the traveller flow swapping from route 1 on day t to route 2 on day t +1 is an increasing function of both:

(1) the flow $X_1(t)$ on the more expensive route 1 on day t; and

(2) the difference $C_1(\mathbf{X}(t)) - C_2(\mathbf{X}(t))$ in route costs on day t.

Here $\mathbf{X}(t) = [X_1(t), X_2(t)]$ is the route flow vector on day t and $C(\mathbf{X}(t)) = [C_1(\mathbf{X}(t)), C_2(\mathbf{X}(t))]$ is the route-cost vector on day t.

Perhaps the simplest swapping hypothesis obeying (1) and (2) above is that the traveller flow swapping from route 1 to route 2 will be proportional to the product of the two factors above in (1) and (2). This "proportional to the product" assumption means, in our day to day context, that for some constant k > 0, the changes $U_1(\mathbf{X}(t)), U_2(\mathbf{X}(t))$ in the traveller flows on routes 1 and 2 will be given by the formulae:

$$U_1(\mathbf{X}(t)) = -kX_1(t)[C_1(\mathbf{X}(t)) - C_2(\mathbf{X}(t))] \tag{2.1}$$

and

$$U_2(\mathbf{X}(t)) = +kX_1(t)[C_1(\mathbf{X}(t)) - C_2(\mathbf{X}(t))]. \tag{2.2}$$

If k is too large (2.1) and (2.2) might give rise to negative flows on day t + 1. So here we are thinking of k as being small – possibly very small. These formulae (2.1) and (2.2) depend on knowing that $C_1(\mathbf{X}(t)) - C_2(\mathbf{X}(t)) > 0$. To make equations (2.1) and (2.2) independent of this knowledge we define (for each real number x):

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

which applies throughout the paper. Using this notation we may write equations (2.1), (2.2) as follows:

$$U_1(\mathbf{X}(t)) = -kX_1(t)[C_1(\mathbf{X}(t)) - C_2(\mathbf{X}(t))]_+ + kX_2(t)[C_2(\mathbf{X}(t)) - C_1(\mathbf{X}(t))]_+ \tag{2.3}$$

and

$$U_2(\mathbf{X}(t)) = +kX_1(t)[C_1(\mathbf{X}(t)) - C_2(\mathbf{X}(t))]_+ - kX_2(t)[C_2(\mathbf{X}(t)) - C_1(\mathbf{X}(t))]_+ \tag{2.4}$$

Given (2.3) and (2.4), our simplest reasonable day-to-day dynamical system becomes:

$$\mathbf{X}(0) = \mathbf{X}^0 \text{ and}$$

$$X_1(t+1) = X_1(t) + U_1(\mathbf{X}(t)) \quad \text{and} \quad X_2(t+1) = X_2(t) + U_2(\mathbf{X}(t))$$

or

$$\mathbf{X}(0) = \mathbf{X}^0 \text{ and}$$

$$\mathbf{X}(t+1) = \mathbf{X}(t) + \mathbf{U}(\mathbf{X}(t)) \tag{2.5}$$

Here $\mathbf{X}(0) = \mathbf{X}^0$ is the starting point (day), and t = 0, 1, 2, 3, ... represents the day-to-day evolution.

To avoid the possibility of negativity here we could utilise a projection in (2.5) but we choose not to do this here. Here we suppose k is very small; aiming to ensure that

$$\mathbf{X}(t+1) = \mathbf{X}(t) + \mathbf{U}(\mathbf{X}(t)) \geq \mathbf{0} \text{ for all t = 1, 2, 3, ...}$$

Let $\Delta_{12} = [-1,1] = (-1,1)^T$ be the swap from route 1 to route 2 vector, and let $\Delta_{21} = [1,-1] = (1,-1)^T$ be the swap from route 2 to route 1 vector. Using these swap vectors we may (instead of using (2.1) – (2.4)) define **U** by:

$$\mathbf{U} = \mathbf{U}(\mathbf{X}) = k\{X_1[C_1(\mathbf{X}) - C_2(\mathbf{X})]_+\Delta_{12} + X_2[C_2(\mathbf{X}) - C_1(\mathbf{X})]_+\Delta_{21}\} \tag{2.6}$$

We have here defined the swap rate $\mathbf{U}(\mathbf{X})$ just between two routes in the simple network in figure 1. But it is easy to generalise and hypothesise that (2.6) might represent a swap rate between any two pairs of routes joining the same OD pair in a general network.

So in a general network let r ~ s initially means that "route r and route s join the same OD pair and are different". Then to generalise (2.6) so that it applies to swaps between all suitable pairs of routes we suppose that there are N routes and specify the co-ordinates Δ_{rsq} of the N-vector Δ_{rs} as follows:

$$\begin{aligned} &\Delta_{rsr} = -1 \text{ if } r \sim s \text{ and } r \neq s, \\ &\Delta_{rss} = +1 \text{ if } r \sim s \text{ and } r \neq s, \\ &\Delta_{rsq} = 0 \quad \text{otherwise.} \end{aligned} \tag{2.7}$$

This vector Δ_{rs} is the swap from route r to route s vector. Then, using (2.7), the general form of (2.6) becomes:

$$\mathbf{U}(\mathbf{X}) = k \sum_{\{(r,s); r \sim s\}} X_r[C_r(\mathbf{X}) - C_s(\mathbf{X})]_+\Delta_{rs} \tag{2.8}$$

This now applies to a network with several OD pairs and N routes and is the PAP direction in Smith (1984a); here in (2.8) swaps between any pair of routes joining the same OD pair are allowed, and all individual terms or swap rates are proportional to flows and cost differences. As before we here suppose that k is very small, so that for any feasible **X (≥ 0), X + U(X)** ≥ **0.**

2.2 Segments, routes and the restricted proportional adjustment process (RPAP)

In this section, we take account of the criticism of He et al (2010) by making " ~ " more restrictive than (2.7) above; we do this by adding the Paired Alternative Segment (PAS) restriction. The PAS restriction further confines the pairs of routes which can occur in the sum (2.8). From now on the only pairs (r, s) which occur in (2.8) must not only join the same OD pair but must also satisfy the PAS restriction.

Paired alternative segments were first introduced in Bar-Gera (2010). Before describing the PAS restriction, we define segments and routes in a general network.

Suppose given a network comprising a set N of nodes and a set L of directed links, where each link a in L has an upstream node and a (distinct) downstream node in N. Consider a non-empty finite connected ordered sequence of distinct links in L and distinct nodes in N:

$$a_1, n_1, a_2, n_2, a_3, , \ldots . , n_{k-2}, a_{k-1}, n_{k-1}, a_k,.$$

Here for each i, node n_i is the downstream node of link a_i and the upstream node of link a_{i+1}.

Definition 1: Such an ordered sequence of links and nodes is called a segment if n_0, the upstream node of link a_1, differs from n_k, the downstream node of link a_k. (So a segment is not a loop.)

The above segment is said to join nodes n_0 and n_k. A segment is connected, has no loops and joins two distinct nodes.

Definition 2: A route is then defined to be a segment. In this paper routes are segments and segments are routes. So routes here have no loops.

We may now define the PAS restriction. Consider two routes R_r and R_s. Let the difference set (or sequence) $R_r – R_s$ be the ordered sequence of all those links and nodes which form part of route R_r but which do not form part of route R_s. This sequence may or may not be a segment because it may have two or more components which are not connected to each other.

Henceforth in this paper, we specify " ~ " in terms of the above definitions as follows. Given two routes R_r and R_s we now re-define r ~ s to mean that:

(a) routes R_r and R_s join the same pair of nodes (as in PAP);

(b) the difference set $R_r – R_s$ is a segment; and

(c) the difference set $R_s – R_r$ is a segment.

(a) - (c) above imply that $R_r – R_s$ and $R_s – R_r$ form a pair of alternative segments: they join the same two nodes (and do not intersect). This concept, of "a pair of alternative segments", was introduced by Bar-Gera (2010).

In this paper, from this point onwards, each term in the sum (2.8) must correspond to a pair (r, s) for which r ~ s in this new sense; which means that each such (r, s) satisfies (a), (b) and (c) above; with this added PAS restriction in the definition of r ~ s, the sum (2.8), of course, now has fewer terms and takes on a new meaning.

This simple change in the definition of ~ in (2.8) makes the flow-swapping much more natural and realistic, because this change removes the anomaly highlighted by He et. al. (2010); it seems likely that this is the simplest way to correct the original PAP for this anomaly.

We may now utilise (2.8) (with the new definition of ~) and also include t (recovering the dynamical system (2.5) as a special case with just two routes) by putting:

$$\mathbf{X}(0) = \mathbf{X}^0$$
$$\mathbf{X}(t+1) = \mathbf{X}(t) + \mathbf{U}(\mathbf{X}(t)) \text{ for } t = 0,1,2,3,.... \tag{2.9}$$

where $\mathbf{X}^0$ is a given supply-feasible starting route flow vector meeting a given demand.

With the above more restrictive definition of r ~ s, equations (2.7), (2.8) and (2.9) together define a restricted proportional adjustment process or an RPAP.

It is possible that the sequence exits the feasible region and then the whole sequence is not well-defined. We address these issues below.

2.3 RPAP in a general network

In this section, we present RPAP in more detail for a general network and show that, under certain conditions, for any given feasible start point $\mathbf{X}(0) = \mathbf{X}^0$ a parameter value k in (2.8) may be found which ensures that the dynamical system (2.7) - (2.8) - (2.9) is well defined.

Definition 3: (Definition of a general capacitated network with a fixed and rigid non-negative demand, and a continuous route cost function.) Here in this paper a general capacitated network is to comprise:

(A1) a standard network (comprising N routes joining K OD pairs) with a route-link incidence matrix **A**;

(A2) for each link i, there is a positive saturation flow s_i defined for all i (S will denote all vectors in R^N such that $\mathbf{AX} < \mathbf{s}$, where $\mathbf{s}$ is the vector of all the s_i);

(A3) for each OD pair p, there is a given fixed (or rigid) non-negative demand ρ_p from the origin node to the destination node (D will denote the set of route-flow N-vectors meeting all these demands which have all co-ordinates non-negative); and

(A4) for each link i there is a non-negative continuous link cost function c_i giving the link cost in terms of the link flow x_i, defined for all $x_i < s_i$ and tending to infinity as $x_i \rightarrow s_i$ from below.

Given a general network as specified above in (A1 – A4), let rjoinsp mean that "route r joins OD pair p". It follows from (2.8) that, for each t, each **X**(t) belonging to D∩S and each OD pair p , the sum of all route-flow swaps is zero at all times, i.e.: for all p,

$$\sum_{\{r;rjoinsp\}} U_r(\mathbf{X}(t)) = 0 \qquad \text{for all t.}$$

Hence, with our new PAS-restricted meaning of ~ in section 2.2, each total OD flow is conserved in the dynamical system (2.7, 2.8, 2.9) or:

$$\sum_{\{r;rjoinsp\}} X_r(t+1) = \sum_{\{r;rjoinsp\}} X_r(t) = \sum_{\{r;rjoinsp\}} X_r(0) = \rho_p \text{ for all } t = 0,1,2,3,...$$

So if the dynamical system (2.7, 2.8, 2.9) starts at $\mathbf{X}(0) = \mathbf{X}^0$ within the set

$$D_1 = \{\mathbf{X}: \sum_{\{r;rjoinsp\}} X_r = \rho_p \text{ for all p}\}$$

then it remains within that set. The demand feasible set is of course further restricted by a non-negativity constraint; so let:

$$D = \{\mathbf{X}: \sum_{\{r;rjoinsp\}} X_r = \rho_p \text{ for all p and } X_s \geq 0 \text{ for all s}\}.$$

Here this set D will be the set of demand-feasible route flow vectors.

Now **A** denotes the link-route incidence matrix and **AX** is the link flow vector corresponding to the route flow vector $\mathbf{X} \geq \mathbf{0}$. The supply feasible set is the set of route flow vectors with all co-ordinates non-negative and which also belong to S where

S = {**X**; **AX** < **s**}.

To establish reasonable values for k in (2.8), initially we suppose that all demand-feasible route flow vectors are also supply-feasible; and that the cost function **C** is thus defined for all **X** in D. Since **C** is in this case a continuous cost function defined on the whole demand feasible set D, **C** must also be bounded on D and for each route r the least upper bound of C_r(**X**) (as **X** varies over D) must be attained. Let M be the maximum of all the route costs C_r(**X**) as **X** varies over D and as r varies (i.e. $1 \leq r \leq N$). We show below that if the value k is chosen such that $k \leq 1/(NM)$, then the dynamical system (2.7, 2.8, 2.9) is well defined in this initial case.

For any $k \leq 1/(NM)$, any demand feasible route flow vector **X** and any route r:

$$\begin{aligned}
[\mathbf{X}+\mathbf{U}(\mathbf{X})]_r &\geq X_r - \sum_{\{s;s\sim r\}} kX_r(C_r - C_s)_+ = X_r[1 - \sum_{\{s;s\sim r\}} k(C_r - C_s)_+] \\
&\geq X_r[1 - \sum_{\{s;s\sim r\}} (C_r - C_s)_+ / (NM)] \geq X_r[1 - \sum_s C_r / (NM)] \\
&\geq X_r[1 - \sum_s M / (NM)] \geq X_r[1 - NM / (NM)] = 0
\end{aligned}$$

Thus under our assumption that **C(X)** is continuous on D and so bounded on D, choosing $k \le 1/(NM)$ will ensure that:

$$\mathbf{X} \ge 0 \Rightarrow \mathbf{X} + \mathbf{U}(\mathbf{X}) \ge 0 .$$

Thus, in this essentially uncapacitated case, if k is small enough

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

and the dynamical system (2.7, 2.8, 2.9) is well defined and yields an infinite sequence of route flow vectors in D.

In the following sub-section 2.4, we suppose that **U(X)** is defined as in (2.7, 2.8) (with the more restricted PAS-restricted meaning of ~) and with $k \le 1/(NM)$. Section 2.4 briefly considers the case where **s** is a large vector. Section 2.5 considers the case where **s** is not large and so may constrain flows.

2.4. A Lyapunov stability result for RPAP when s is large and route-costs are bounded on D.

In this section, we discuss the stability of the RPAP dynamical route choice model for the general network defined in Section 2.3 when the route cost function **C = C(X)** is defined and bounded on the whole demand-feasible set D.

Firstly, we specify a measure of dis-equilibrium. Following Smith (1984a), and bearing in mind the PAS modification we have now added by restricting the pairs (r, s) for which r ~ s, we define the PAS-modified objective function V as:

$$\mathrm{V}(\mathbf{X}) = \sum_{\{(r,s);r\sim s\}} \mathrm{X}_r[\mathrm{C}_r(\mathbf{X}) - \mathrm{C}_s(\mathbf{X})]_+^2 \qquad \text{for all } \mathbf{X} \in \mathrm{D}. \tag{2.10}$$

Then V is a measure of departure from equilibrium. It is easy to see here that for $\mathbf{X} \in$ D:

$\mathrm{V}(\mathbf{X}) = 0$ if and only if {for all r, s such that r ~ s, $\mathrm{X}_r[\mathrm{C}_r(\mathbf{X}) - \mathrm{C}_s(\mathbf{X})]_+^2 = 0$ };

if and only if {for all r, s such that r ~ s, $[\mathrm{C}_r(\mathbf{X}) - \mathrm{C}_s(\mathbf{X})] > 0 \Rightarrow \mathrm{X}_r = 0$};

if and only if $\mathbf{X}$ is a Wardrop equilibrium (Wardrop, 1952).

The set E of Wardrop equilibria may thus be specified as follows:

$$\mathrm{E} = \{\mathbf{X} \in \mathrm{D}; \mathrm{V}(\mathbf{X}) = 0\}.$$

It is natural to consider approximate equilibria, so let (for any $\varepsilon > 0$),

$$\mathrm{E}_\varepsilon = \{\mathbf{X} \in \mathrm{D}; \mathrm{V}(\mathbf{X}) \le \varepsilon\} \tag{2.11}$$

The following lemma and theorem prove a stability result for the dynamical system (2.7, 2.8, 2.9) when the route cost function is bounded on D.

Lemma 1. Suppose that $\mathbf{C}$ is monotone and directionally differentiable in all feasible directions throughout D. Suppose also that the directional derivative **C'**($\mathbf{X}$; $\boldsymbol{\delta}$**)** is a continuous function of ($\mathbf{X}$, $\boldsymbol{\delta}$). Then

(a) V is directionally differentiable in all feasible directions at each $\mathbf{X}$ in D;

(b) $\mathrm{V}'(\mathbf{X};\mathbf{U}(\mathbf{X}))$ is a continuous function of $\mathbf{X}$ in D; and

(c) $\mathrm{V}'(\mathbf{X};\mathbf{U}(\mathbf{X}))<0$ for all non-equilibrium $\mathbf{X}$ in D.

PROOF. See Appendix B.1

It follows from Lemma 1 that

$$\mathrm{V}'(\mathbf{X};\mathbf{U}(\mathbf{X}))=\lim_{h\to 0+}[\mathrm{V}(\mathbf{X}+h\mathbf{U}(\mathbf{X}))-\mathrm{V}(\mathbf{X})]/h<0 \tag{2.12}$$

for all non-equilibrium $\mathbf{X}\in\mathrm{D}$. **(**Here h represents a small step-length. $\mathrm{V}'(\mathbf{X};\mathbf{U}(\mathbf{X}))$ as defined in (2.12) is the directional derivative of V at $\mathbf{X}$ in the direction $\mathbf{U}(\mathbf{X})$). Furthermore, part (c) of lemma 1 (see appendix B1) shows that for any non-equilibrium $\mathbf{X}\in\mathrm{D}$,

$$\mathrm{V}'(\mathbf{X};\mathbf{U}(\mathbf{X}))<-\sum_{\{(r,s);r\sim s\}}\mathrm{X}_r\{[\mathrm{C}_r(\mathbf{X})-\mathrm{C}_s(\mathbf{X})]_+\}^3 \tag{2.13}$$

Of course $\sum_{\{(r,s);r\sim s\}}\mathrm{X}_r\{[\mathrm{C}_r(\mathbf{X})-\mathrm{C}_s(\mathbf{X})]_+\}^3$ is positive away from equilibrium and so (2.13) implies that, away from equilibrium, (2.12) holds and that V is a reasonable Lyapunov function (Lyapunov, 1907) for the dynamical system (2.7) + (2.8) + (2.9).

It may then be shown (by following Smith (1984a) but using the RPAP-restricted swaps here) that, under reasonable additional conditions, following $\mathbf{U}(\mathbf{X})$ in a smooth model causes $\mathrm{V}(\mathbf{X})$ to converge to zero and so causes $\mathbf{X}$ to approach the equilibrium set $\mathrm{E}=\{\mathbf{X};\mathrm{V}(\mathbf{X})=0\}$ as time passes. It may further be shown that provided k is chosen as above and then is further chosen to be small the dynamical system (2.9) reaches the set of approximate equilibria. (This further result is contained in the result proved in section 2.5 below which is stronger since **s** is taken into account.)

2.5 A Lyapunov stability result for RPAP when s is not large and route-costs are not bounded on D

In this section, we demonstrate a Lyapunov stability result for a route-flow swapping dynamical system based on RPAP when **s** is not large and so the saturation flows may (potentially) actively restrict link flows. In order to do this we need to slightly extend the result proved in section 2.3 to allow for the upper bounds on link flows.

The extended non-negativity result. This is as follows. Let $\mathbf{X}^0 \in$ D and $\mathbf{AX}^0 < \mathbf{s}$; so that $\mathbf{X}^0$ is both demand and supply feasible. Then there is k > 0 (depending on $\mathbf{X}^0$) such that

$$\text{if } \mathbf{X} \in \text{D}, \mathbf{AX} < \mathbf{s} \text{ and } V(\mathbf{X}) \leq V(\mathbf{X}^0) \text{ then } \mathbf{X} + \mathbf{U}(\mathbf{X}) \geq \mathbf{0}.$$

Here, in what follows in theorem 1, $\mathbf{U}(\mathbf{X}) = \mathbf{U}_k(\mathbf{X})$ is defined in equation (2.8), using the above k. The proof of this extended non-negativity result is a simple extension of the proof in section 2.3 to take account of the upper limits on link flows; we do not give this extension here as it is straightforward.

We will also need here lemmas 2-4 below.

Lemma 2. In a standard capacitated network (see A1 – A4 above in definition 3) with demand D, let $\mathbf{X}^0 \in$ D and $\mathbf{AX}^0 < \mathbf{s}$ (so that $\mathbf{X}^0$ is both supply-feasible and demand feasible). Also let $\mathbf{X^*} \in$ D where $\mathbf{AX}^* \leq \mathbf{s}$ but $\mathbf{AX}^* < \mathbf{s}$ no longer holds (so $\mathbf{X^*}$ is demand feasible, not supply-feasible and very nearly supply-feasible). Let $\mathbf{X}$ move along the straight line joining $\mathbf{X}^0$ and $\mathbf{X^*}$. Then $V(\mathbf{X})$ tends to $+\infty$ *as* $\mathbf{X} \rightarrow \mathbf{X}^*$.

PROOF. See Appendix B.2.

Lemmas 3 and 4 below show that monotonicity and directional differentiability of the route cost function **C** follow from corresponding properties of the link cost function **c**.

Lemma 3. Suppose that the link cost function $\mathbf{c} = \mathbf{c}(\mathbf{x})$ is monotone, then the route cost function $\mathbf{C} = \mathbf{C}(\mathbf{X})$ is also monotone.

PROOF. See Appendix B.3.

Lemma 4. Suppose that the link cost function $\mathbf{c} = \mathbf{c}(\mathbf{x})$ is directionally differentiable in all feasible directions, then the route cost function $\mathbf{C} = \mathbf{C}(\mathbf{X})$ is also directionally differentiable in all feasible directions. Further the directional derivative of **C** is continuous if the directional derivative of **c** is continuous.

PROOF. See Appendix B.4.

Short statement of theorem 1: Given a feasible start point $\mathbf{X}^0$, k satisfying the extended non-negativity result above and a non-empty set E_ε of approximate equilibria; under reasonable conditions there is h > 0 such that the dynamical system $\mathbf{X} \rightarrow \mathbf{X} + h\mathbf{U}(\mathbf{X})$ (starting at $\mathbf{X}^0$) enters E_ε in a finite number of steps. Here all links are capacitated. (Note that h in general depends on the starting point $\mathbf{X}^0$.)

Theorem 1

Suppose given:

(1a) a network comprising N routes joining K OD pairs;

(1b) for each of the K (origin node, destination node) pairs a fixed or rigid non-negative demand from the origin node to the destination node (D denotes the set of route-flow N-vectors **X** meeting these given demands); and

(1c) for each link a a positive saturation flow s_a and a continuous, non-negative and non-decreasing continuously differentiable cost function $c_a(.)$, defined for all link flows $x_a < s_a$ and tending to $+\infty$ as $x_a \rightarrow s_a$.

Suppose further that the set $S \cap D$ is non-empty and for any $\varepsilon > 0$ let

$$E_\varepsilon = \{\mathbf{Y} \in S \cap D;\ V(\mathbf{Y}) \le \varepsilon\}$$

be a given set of approximate equilibria, where V is given in (2.10). Let $\mathbf{X}^0 \in S \cap D$ be any feasible starting route flow vector, and for this starting route flow vector let

$$D^{0\varepsilon} = \{\mathbf{Y} \in S \cap D;\ \varepsilon \le V(\mathbf{Y}) \le V(\mathbf{X}^0)\}$$

so that

$$D^{00} = \{\mathbf{Y} \in S \cap D;\ 0 \le V(\mathbf{Y}) \le V(\mathbf{X}^0)\}.$$

Then:

(i) given the start point $\mathbf{X}^0 \in S \cap D$, given any k satisfying the above extended non-negativity condition, and given $\mathbf{U}(\mathbf{X})$ determined by equation (2.8), there is $h_0 > 0$ but so small that:

$$\mathbf{X} \in D^{00} \Rightarrow \mathbf{X} + h\mathbf{U}(\mathbf{X}) \in D^{00} = \{\mathbf{Y} \in S \cap D;\ 0 \le V(\mathbf{Y}) \le V(\mathbf{X}^0)\}.$$

for all h such that $0 < h \le h_0$.

Now, given the start point $\mathbf{X}^0 \in S \cap D$ and the particular h_0 constructed in (i), let $0 < h < h_0$ so and define $T = T_h: D^{00} \rightarrow D^{00}$ as follows:

$$T(\mathbf{X}) = \mathbf{X} + h\mathbf{U}(\mathbf{X}) \tag{2.14}$$

for all $\mathbf{X} \in D^{00}$. Consider (for h satisfying $0 < h < h_0$) the (well-defined) sequence

$$\mathbf{X}^0, T\mathbf{X}^0, T^2\mathbf{X}^0, \ldots, T^{n-2}\mathbf{X}^0, T^{n-1}\mathbf{X}^0, T^n\mathbf{X}^0, \ldots . \tag{2.15}$$

Then

(ii): the infinite sequence (2.15) enters E_ε at some value of n.

PROOF. See Appendix B. 5

The above result depends on monotonicity and directional differentiability of the route cost function. It proves that under reasonable conditions the dynamical model of route-flow adjustment given in (2.14) and (2.15), based on RPAP, eventually reaches a set of approximate equilibria. This limited convergence result may be strengthened to show that there is an $h > 0$ such that (2.15) enters the set E_ε and remains within that set for all remaining time.

3. Extending RPAP to Embrace Signal Control Adjustment Using "Red-Time" Costs.

In this section, we consider dynamics as both route-flows and signal green-times vary. We show how certain signal control policies may easily be included within the above RPAP route adjustment process (2.8) + (2.9); and that if the special control policy P_0 is included in this way then a convergence result very similar to the "no control" Theorem 1 holds.

First it is necessary to outline a method which makes signal green-times responsive to "current" flows and delays. Normally this is done (in practice) by using signal stages and utilizing a signal control policy stating how stage green times vary with traffic flow. (In this paper a stage is a (maximal) set of links terminating at a single junction which has the property that when the stage is green then all links in that stage are green.)

Here we specify a corresponding procedure by using red-times, "antistages" and red-time costs following Smith and Mounce (2011); this leads to a two-commodity link flow model.

In this paper, a stage (or a link) green-time is the proportion of time that the stage (or the link) is green. A link green-time is obtained by adding relevant stage green-tmes. Furthermore, we assume that there are no minimum green times and that if a link is green, then all movements leaving the link are given green. (These are control idealisations employed in this paper.)

3.1 Anti-stages and red-time costs

Given any signal stage, say stage J (at a certain junction) there is a corresponding antistage AJ: comprising all those links (terminating at the same junction) which are not in stage J. Antistages have the property that when an antistage is red then all links in that antistage are red. Now to specify a signal control policy, instead of specifying how the green-times allocated to the signal stages vary with link traffic flows, we here specify how the red times allocated to the anti-stages vary with link traffic flows. Any responsive control policy stated in terms of green times and stages may be written in terms of red times and antistages. Link red times are sums of antistage red times just as link green times are sums of stage green times.

An extended version of the route-swapping dynamical system (2.8) – (2.9) is then constructed by thinking of link red-time as an extra "flow" through each signal controlled link exit, causing an extra cost. In this extended system, link costs add to give route costs, for each pair of routes real traffic flow swaps toward the cheaper route; also link red-costs add to give antistage costs and, for each pair of antistages, red-time swaps toward the "cheaper" antistages. The route flow swaps and the antistage red-time swaps both follow the same proportional rule described in Section 2.3 for just flows.

The aggregated flow on link i will comprise the flow of real vehicles added to a suitable multiple of link i "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 aggregated "flow volume" be $x_i + s_i r_i$; where x_i represents the "real" vehicular flow and r_i represents the proportion of time approach i is red. The multiple $s_i r_i$ is the capacity lost due to the proportion (r_i) of red time, bearing in mind the saturation flow s_i at the link exit. Then we suppose that the cost (or travel time) of traversing approach i equals

$$c_i(x_i) + b_i(x_i + s_i r_i)\,. \tag{3.1}$$

Here $c_i(x_i)$ represents the cost of traversing the length of the link when the flow is x_i and $b_i(x_i + s_i r_i)$ represents the bottleneck delay felt at the traffic signal when the flow is x_i and the red time proportion is r_i. Both $c_i(.)$ and $b_i(.)$ are non-decreasing real-valued functions of a real variable. Here the gradient of c_i may be rather small and the gradient of b_i may be rather large: b_i may even have a vertical asymptote at s_i; in fact below we suppose that this is so.

3.2 A two-commodity link model and RPAP dynamics of (route-flow, red-time) vectors

This approach (using (3.1)) allows a simple dynamical model of control and routeing to be constructed. In essence we have a two-commodity link model where the two commodities are:

x_i = vehicular flow on link i and

r_i = red-time on link i (a proportion and dimensionless).

The dynamics are now to be as follows (for any responsive control policy). At each origin real flow switches to cheaper routes following RPAP in Section 2.3, pushed by sums of the "standard" link costs $c_i(x_i) + b_i(x_i + s_i r_i)$; but now also red-time switches to "cheaper" antistages pushed by antistage "red-time-costs"; again still essentially following RPAP, although with the antistage red-time adjustment there is no distinction between PAP and RPAP.

The link red-time-costs are defined to suit, or to define, a particular signal control policy. For example, the equi-saturation policy (Webster, 1957) may be obtained if:

$$\text{the link red-time cost} = \frac{x_i}{g_i s_i} = \frac{x_i}{(1-r_i)s_i} = \frac{x_i}{(s_i - r_i s_i)}. \qquad (3.2a)$$

The P_0 policy (Smith, 1979a, b, c) is obtained if:

$$\text{the link red time cost} = s_i b_i(x_i + s_i r_i). \qquad (3.2b)$$

For example, for a signalised junction with two approaches, the above red-time swapping specifications implied in (3.2a, b) may be thought of as having the following two objectives:

(i) The equi-saturation policy chooses green times which seek to ensure that:

$$\frac{x_1}{g_1 s_1} = \frac{x_2}{g_2 s_2} \quad \text{or} \quad \frac{x_1}{(1-r_1)s_1} = \frac{x_2}{(1-r_2)s_2} \qquad (3.3a)$$

as if this holds then no more red-time swapping occurs.

(ii) The P_0 policy chooses red times which seek to ensure that

$$s_1 b_1(x_1 + s_1 r_1) = s_2 b_2(x_2 + s_2 r_2); \qquad (3.3b)$$

if this holds then no more red-time swapping occurs.

It is clear from the above equations (3.2b) and (3.3b) 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 the red-time vector $\mathbf{r}$) seek to ensure that the bottleneck delay b_2 will tend to be small; encouraging the use of the approach with the higher saturation flow (even if the actual flow on that approach is small). The policy encourages re-routeing toward higher capacity routes rather than rewarding travellers on existing routes. It may be shown that, under natural conditions, which include strict capacity restrictions, this policy maximises network throughput at an equilibrium distribution of traffic flows. Theorem 2 below may be regarded as a simple demonstration of this.

3.3 RPAP for the combined dynamical system with the P_0 policy in a general network

In this section, we extend the previous Theorem 1 above so as to include the responsive policy P_0 within the day-to-day dynamic RPAP framework for a general network.

Suppose that the P_0 signal control policy is employed at each node of a network. The 2-commodity link i cost-flow function arising is then, following the previous section:

$$[c_i(x_i) + b_i(x_i + s_i r_i), s_i b_i(x_i + s_i r_i)]. \tag{3.4}$$

(Lemma 5 below shows that (3.4) is monotone if c_i and b_i are both monotone.)

The first co-ordinate in (3.4) gives the link cost felt by real flow on link i and the second co-ordinate gives the red-cost felt by red-times. We need to extend (2.9) so as to include the red times of antistages as well as the flows along routes. In doing this we think of signal antistages as new "routes" and red-times as new "flows" on those new routes, and apply dynamics like (2.8) to both.

Both the route-flow switches and the stage-red-time switches will depend, in essentially the same way, on the specifications of costs of routes and antistages, which are both sets of links. These costs are determined as follows:

(a) for each route r : the link i costs $c_i(x_i) + b_i(x_i + s_i r_i)$ are added over all links i in route r to determine route r (flow-)cost, and

(b) for each antistage J : the link i red-time costs $s_i b_i(x_i + s_i r_i)$ are added over all links in antistage J to determine antistage J (red-time-)cost.

Consider a general network. Suppose that $\mathbf{A}$ is the link-route incidence matrix and that $\mathbf{B}$ is the link-antistage incidence matrix. Let $\mathbf{X}$ be a vector of route flows and let $\mathbf{R}$ be a vector of antistage red-times. Then the link flow vector $\mathbf{x} = \mathbf{AX}$ and the link red-time vector $\mathbf{r} = \mathbf{BR}$.

The two-commodity demand set is now $D \times RD$ where D is the set of demand feasible route-flow vectors and RD is the set of demand-feasible anti-stage red-time vectors.

Then the set S' which guarantees supply-feasibility for this two-commodity network is now defined as follows:

$$S' = \{(\mathbf{X}, \mathbf{R}); \mathbf{AX} + \mathbf{s} \bullet \mathbf{BR} < \mathbf{s}\} = \{(\mathbf{X}, \mathbf{R}); (\mathbf{AX})_i + s_i (\mathbf{BR})_i < s_i \quad \text{for all } i = 1, 2, 3, ..., n\}$$

For all non-negative vectors $(\mathbf{X}, \mathbf{R}) \in S'$, the flow cost $C_r = C_r(\mathbf{X}, \mathbf{R})$ of route r depends on two commodities and is to be given by:

$$C_r(\mathbf{X}, \mathbf{R}) = \sum_{i \sim r} [c_i(x_i) + b_i(x_i + s_i r_i)] \tag{3.5}$$

where $i \sim r$ if link i forms part of route r; and (as we are using P_0) the red-time-cost of antistage A_J

$$AC_J = AC_J(\mathbf{X}, \mathbf{R}) = \sum_{i \sim AJ} [s_i b_i (x_i + s_i r_i)] \tag{3.6}$$

where $i \sim AJ$ means that link i is in antistage J. The formula here for the antistage cost AC_J arises from the definition of the P_0 signal control policy in Smith (1979a) and the simple link delay formulation $b_i(x_i + s_i r_i)$ adopted in this paper.

In lemma 5 below we drop suffices and assume that the link i cost function c and the link i bottleneck function b are both non-decreasing so that they are both monotone one-dimensional functions.

Lemma 5. Suppose that c is monotone and b is monotone. Then $[c(x) + b(x + sr), sb(x + sr)]$ is a monotone function of the 2-vector (x, r).

PROOF. See Appendix B.6.

It follows from Lemma 5 that (3.4) is monotone and then it follows (essentially from the two-commodity versions of lemmas 3 and 4 above) that

[route cost, antistage red-time cost] = $[\mathbf{C}(\mathbf{X}, \mathbf{R}), \mathbf{AC}(\mathbf{X}, \mathbf{R})]$

is a monotone directionally differentiable function of $[\mathbf{X}, \mathbf{R}]$ throughout $S' \cap (D \times RD)$.

The combined (route flow, red-time) adjustment direction is now defined to be

$$\mathbf{U}(\mathbf{X}, \mathbf{R}) = [\mathbf{U}^{\mathbf{X}}(\mathbf{X}, \mathbf{R}), \mathbf{U}^{R}(\mathbf{X}, \mathbf{R})]$$

where

$$\mathbf{U}^{\mathbf{X}}(\mathbf{X}, \mathbf{R}) = k \sum_{\{(r,s): r \sim s\}} X_r [C_r(\mathbf{X}, \mathbf{R}) - C_s(\mathbf{X}, \mathbf{R})]_+ \Delta_{rs} \tag{3.7}$$

and

$$\mathbf{U^R}(\mathbf{X},\mathbf{R}) = k \sum_{\{(I,J):I\sim J\}} R_I [AC_I(\mathbf{X},\mathbf{R}) - AC_J(\mathbf{X},\mathbf{R})]_+ \delta_{IJ} \tag{3.8}$$

Here k is to be chosen so that (**X**, **R**) + **U**(**X**, **R**) ≥ **0** for relevant feasible **(X, R),** and for two antistages I and J, I ~ J if and only if these antistages are (a) different and (b) at the same junction (so that swapping red-time between them is sensible). The antistage red-time swap vector $\boldsymbol{\delta}_{IJ}$ has -1 in the Ith place and +1 in the Jth place and zeros everywhere else. Moving **R** in direction $\boldsymbol{\delta}_{IJ}$ swaps red-time from antistage I to antistage J.

Extending the objective function V given in (2.10), to allow for the current context involving red-time swaps as well as route-flow swaps, let

$$V(\mathbf{X},\mathbf{R}) = \sum_{\{(r,s):r\sim s\}} X_r [C_r(\mathbf{X},\mathbf{R}) - C_s(\mathbf{X},\mathbf{R})]_+^2 + \sum_{\{(I,J):I\sim J\}} R_I [AC_I(\mathbf{X},\mathbf{R}) - AC_J(\mathbf{X},\mathbf{R})]_+^2 \tag{3.9}$$

This extended V is defined throughout S' ∩ (D×RD) and will be the Lyapunov function for a dynamical system following directions (3.7), (3.8). Now RD denotes the set of (demand-) feasible antistage red time vectors **R**. Then the set E of consistent equilibria is here specified as follows

$$E = \{[\mathbf{X},\mathbf{R}] \in D \times RD; \quad V([\mathbf{X},\mathbf{R}]) = 0\}.$$

It is natural to consider also approximate consistent equilibria, so let, for any $\varepsilon > 0$,

$$E_\varepsilon = \{[\mathbf{X},\mathbf{R}] \in D \times RD; \quad V([\mathbf{X},\mathbf{R}]) \le \varepsilon\}.$$

It follows from Lemma 5 and by expanding Lemma 1 that $[\mathbf{U^X}(\mathbf{X},\mathbf{R}), \mathbf{U^R}(\mathbf{X},\mathbf{R})]$ is a descent direction for $V(\mathbf{X},\mathbf{R})$ (at any feasible $(\mathbf{X},\mathbf{R})$ which is not an equilibrium consistent with the P_0 control policy). Then a combined (route flow, antistage red-time) dynamical system is (where $[\mathbf{X}^0, \mathbf{R}^0]$ is a feasible starting pair):

$$[\mathbf{X}(t+1), \mathbf{R}(t+1)] = [\mathbf{X}(t), \mathbf{R}(t)] + h[\mathbf{U^X}(\mathbf{X},\mathbf{R})(t), \mathbf{U^R}(\mathbf{X},\mathbf{R})(t)] \text{ for t = 0, 1, 2, 3, …}$$

$$[\mathbf{X}(0), \mathbf{R}(0)] = [\mathbf{X}^0, \mathbf{R}^0] \tag{3.10}$$

where 0 < h ≤ 1. A modification of the flow-only Theorem 1 may now be proved for this [route flow, anti-stage red-time] dynamical system. This is theorem 2 below.

The proof of theorem 2 below depends on noting that lemma 1 holds also in this "two-commodity" setting; the proof of this is straightforward and yields this "two commodity" result

Two commodity form of lemma 1. Suppose that [**C** , **AC**] is monotone and directionally differentiable throughout $S' \cap (D \times RD)$. Suppose also that the directional derivative [**C** , **AC**]'((**X, R**); **δ)** in direction **δ** is a continuous function of ((**X, R**), **δ**), then

(a) V is directionally differentiable in all feasible directions at all **(X, R)** in $S' \cap (D \times RD)$;

(b) V'((**X, R**); **U(X, R)**) is a continuous function of (**X, R**) in $S' \cap (D \times RD)$; and

(c) V'((**X, R**); **U**(**X, R**)) < 0 for all non-equilibrium (**X, R**) in $S' \cap (D \times RD)$.

3.4 Lyapunov stability of the combined system with RPAP and the P_0 control policy

Extending the previous theorem 1 to allow for the two commodities we obtain theorem 2 below.

Short statement of theorem 2: In the following theorem 2 we show that under reasonable conditions the dynamical system (**X**, **R**) → (**X**, **R**) + h**U**(**X**, **R**) (starting at any feasible $(\mathbf{X}, \mathbf{R})^{\mathbf{0}}$) enters a given set of approximately P_0 consistent equilibria in a finite number of steps. Here (as in theorem 1) all links are capacitated. k is supposed chosen to ensure that a two-dimensional version of the extended non-negativity condition in section 2.5 holds; This condition will involve (**X, R**) and **U(X, R)** rather than just **X** and **U(X).**

Theorem 2.

Suppose given:

(2a) a network comprising N_1 routes joining K_1 OD pairs and N_2 antistages at K_2 junctions;

(2b) for each of the K_1 (origin node, destination node) pairs, a fixed or rigid non-negative demand from the origin node to the destination node (D denotes the set of route-flow N_1-vectors **X** meeting these given demands); and for each of the K_2 junctions, there is a set of stages and antistages (RD denotes the set of antistage red-time N_2-vectors **R** arising from all these given antistages); and

(2c) for each link a a positive saturation flow s_a, a continuous, non-negative and non-decreasing continuously differentiable cost function $c_a(.)$, defined for all link flows $x_a \le s_a$ and a continuous non-decreasing continuously differentiable function $b_a(.)$, defined for all link volumes $v_a < s_a$ tending to $+\infty$ as $v_a \to s_a$.

Suppose further that the set $S' \cap (D \times RD)$ is non-empty and for any $\varepsilon > 0$ let

$$E_\varepsilon = \{(\mathbf{X}, \mathbf{R}) \in S' \cap (D\times RD);\ V(\mathbf{X}, \mathbf{R}) \leq \varepsilon\}$$

be a given set of approximate equilibria, where V is given in (3.9). Let $(\mathbf{X}^0, \mathbf{R}^0) = (\mathbf{X}, \mathbf{R})^0 \in S' \cap (D\times RD)$ be any feasible starting (route flow vector, antistage red-time vector) and for this starting (route flow vector, antistage red time vector) $(\mathbf{X}^0, \mathbf{R}^0)$ let

$$(D\times RD)^{0\varepsilon} = \{(\mathbf{X}, \mathbf{R}) \in S' \cap (D\times RD);\ \varepsilon \leq V(\mathbf{X}, \mathbf{R}) \leq V((\mathbf{X}, \mathbf{R})^0)\}.$$

Then:

(i) given the start point $(\mathbf{X}^0, \mathbf{R}^0) = (\mathbf{X}, \mathbf{R})^0 \in S' \cap (D\times RD)$, given any k satisfying the two-dimensional form of the extended non-negativity condition in section 2.5 and given $\mathbf{U}(\mathbf{X}, \mathbf{R})$ determined by equations (3.7) and (3.8), there is $h_0 > 0$ but so small that:

$$(\mathbf{X}, \mathbf{R})\in(D\times RD)^{00} \Rightarrow (\mathbf{X}, \mathbf{R})+h\mathbf{U}(\mathbf{X},\mathbf{R}) \in (D\times RD)^{00} = \{(\mathbf{X}, \mathbf{R})\in S'\cap(D\times RD);\ 0\leq V(\mathbf{X}, \mathbf{R})\leq V((\mathbf{X}, \mathbf{R})^0)\}$$

for all h such that $0 < h \leq h_0$. Here V is given by equation (3.9).

Now, given the start point $(\mathbf{X}^0, \mathbf{R}^0) = (\mathbf{X}, \mathbf{R})^0 \in S' \cap (D\times RD)$ and h_0 satisfying (i) above, let $0 < h \leq h_0$ and define $T = T_h: (D\times RD)^{00} \rightarrow (D\times RD)^{00}$ as follows:

$$T(\mathbf{X}, \mathbf{R}) = (\mathbf{X}, \mathbf{R}) + h\mathbf{U}(\mathbf{X}, \mathbf{R})$$

for all $(\mathbf{X}, \mathbf{R}) \in (D\times RD)^{00}$. Consider (for h satisfying $0 < h < h_0$) the (well-defined) sequence

$$(\mathbf{X}, \mathbf{R})^0, T(\mathbf{X}, \mathbf{R})^0, T^2(\mathbf{X}, \mathbf{R})^0, \ldots, T^{n-2}(\mathbf{X}, \mathbf{R})^0, T^{n-1}(\mathbf{X}, \mathbf{R})^0, T^n(\mathbf{X}, \mathbf{R})^0, \ldots \quad (3.11)$$

Then

(ii) the infinite sequence (3.11) enters E_ε at some value of n.

PROOF. This essentially follows from the proof of Theorem 1, but of course the two commodity version of lemma 1 is needed here.

Theorem 2 assumed that the [flow, antistage] conservation constraints hold at the start point; then using only the not-too-large switches specified via (3.7) and (3.8) ensures that these constraints continue to hold. This is a stability / convergence result where both flow and red-times move simultaneously; and uses the special delay formula $b_i(x_i+s_ir_i)$. It may be interpreted as a rudimentary capacity-maximisation result: at each day the flows on the network satisfy the demand and costs are bounded on the whole sequence.

4. A Simple Example Network: Instability with Equi-saturation and Stability with P_0

In this section we show that rather negative results (instability) arise in some circumstances with the equi-saturation policy and a certain delay formula, whilst stability is maintained with this delay formula with the P0 policy.

Consider the simple network shown in Fig. 2. The network is comprised of a single origin-destination (OD) pair, and two links joining the origin to a signal-controlled junction. The saturation flows on the two approaches to the signalised node are s_1 and s_2, and the free-flow travel times on the two routes are K_1 and K_2 respectively. The total OD flow is T vehicles per minute and the proportions of drivers using each of the two routes are H_1 and H_2 where $H_1 + H_2 = 1$. So the flows on the two routes are $X_1 = TH_1$ and $X_2 = TH_2$ (vehicles per minute).

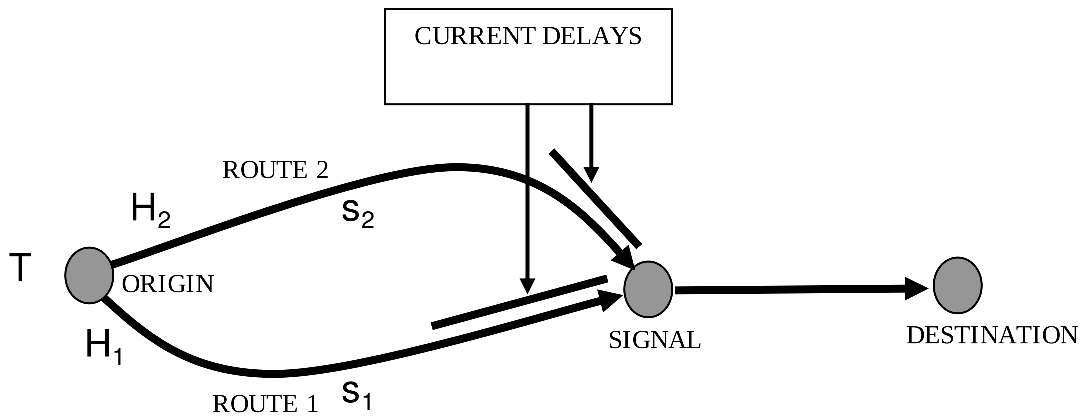

Figure 2. A 3-link 2-route network with a signal-controlled junction. The saturation flows at the junction are s_1 and s_2. The proportion of the total flow rate T using route 1 is H_1 and the proportion of the total flow rate T using route 2 is H_2.

4.1 Route cost functions and link cost functions

We consider a general route cost function having three terms: a free-flow travel time, a rather shallow flow-related travel time, and a rather steep delay function involving both flows and signal green times. This is:

$$C_i = K_i + AX_i + d_i, \tag{4.1}$$

where K_i is the free-flow travel time on route 1 (minutes), X_i is the flow (veh/min) on route i. A is a per-vehicle travel time (min/veh) and is a constant. One form of the steep delay formula is that of Webster's random delay term; which is:

$$d_i = \frac{BX_i}{s_i G_i (s_i G_i - X_i)} \tag{4.2}$$

where G_i is the green-time proportion, s_i is the saturation flow on approach i, and B is a constant. In the above simple case of Figure 2, the flow on approach i is the same as flow on route i.

The delay function of (4.2) is exactly the second term of Webster's delay formula when $B = 9/20$ (Webster, 1958). In Webster's famous two term delay formula, the first term estimates the delay due to the stop-start nature of traffic signal operation (assuming that flow is steady). The second term, used here in (4.2), allows for the random nature of arrivals.

Webster's formula is closely related to the Pollaczek-Khintchine (P-K) formula (Pollaczek, 1930; Khintchin, 1932) for the average waiting time felt by a Poisson stream of arrivals (with arrival rate X vehicles per minute) at a single server (with a constant service rate sG vehicles per minute). This formula is obtained by taking $B = 1/2$ instead of 9/20 in (4.2), (See, for example, Madan and Saleh, 2001).

Now in our network in figure 2, the above delay formula (4.2) may be written:

$$d_i = B\left[\frac{1}{s_i - (X_i + s_i R_i)} - \frac{1}{s_i - s_i R_i}\right] \tag{4.3}$$

The first term of (4.3) is also a non-decreasing function of the red-time proportion to the route R_i and we write it as:

$$w_i = \frac{1}{2}\left[\frac{1}{s_i - (X_i + s_i R_i)}\right], \tag{4.4}$$

We show later that when this delay term w_i is combined with the P0 policy, it yields stable control solutions.

4.2 Flow swapping

It is assumed in this section that:

(1) travellers will stay on the same route if there is no cheaper route.

(2) If there is a route with a smaller expected travel time, then some travellers will periodically swap their route for a quicker route.

By symmetry equal flows on the two routes will yield an equilibrium, so no swapping occurs in that case under the above assumptions (1) and (2). We examine the equilibrium solutions for the asymmetrical flow patterns and see how such asymmetrical flows evolve as time passes allowing for the responsive control; using route-swaps like those specified previously.

In our study here, by computing the times to traverse the two routes using the various delay formula in Section 4.1 when different policies are used to set the signals, the whole triangle of feasible flows was filled with small vectors indicating the direction of motion on the above assumptions (1) and (2).

4.3 Equi-saturation and P_0 policy to the simple network and flow feasibility

For the example network of Figure 2, the equi-saturation policy yields:

$$\frac{X_1}{s_1 G_1} = \frac{X_2}{s_2 G_2}$$

Here $X_1 = TH_1$ and $X_2 = TH_2$, and since $G_1 + G_2 = 1$, we obtain the green-time proportions as:

$$G_1 = \frac{H_1 / s_1}{H_1 / s_1 + H_2 / s_2} \quad \text{and} \quad G_2 = \frac{H_2 / s_2}{H_1 / s_1 + H_2 / s_2}, \tag{4.5}$$

For the P_0 policy, we apply the first term of the P-K wait time function, i.e. (4.4). Then (3.3b) becomes:

$$\frac{Bs_1}{[s_1 - (X_1 + s_1 R_1)]} = \frac{Bs_2}{[s_2 - (X_2 + s_2 R_2)]} \tag{4.6}$$

Together with the condition: $R_1 + R_2 = 1$, we solve (4.6) and obtain the red-time proportions as:

$$R_1 = B[1 - (\frac{X_1}{s_1} - \frac{X_2}{s_2})] \quad \text{and} \quad R_2 = B[1 + (\frac{X_1}{s_1} - \frac{X_2}{s_2})] \tag{4.7}$$

We apply the delay formula (4.1) and (4.2) with the equi-saturation policy. To avoid a zero denominator in (4.2), and to ensure delays given by (4.2) are non-negative, we have:

$$s_1 G_1 - X_1 > 0 \quad \text{and} \quad s_2 G_2 - X_2 > 0.$$

From the above, we obtain the supply-feasibility constraint on the total flow T and flow splits H_1 and H_2 as follows:

$$T < \frac{1}{H_1 / s_1 + H_2 / s_2} \tag{4.8}$$

It can be shown that the second term of the P-K wait time function in (4.3) is feasible also if (4.8) holds. Thus the set of supply-feasible $(T, \mathbf{H})$ pairs is the set of those $(T, \mathbf{H})$ satisfying (4.8). For any given T, (4.8) also defines the supply-feasible set of vectors $\mathbf{H}$. So:

$$S = \{\mathbf{H} : H_1 + H_2, H_1 \geq 0, H_2 \geq 0, \text{and } H_1 \text{ and } H_2 \text{ satisfy inequality (4.8)}\}$$

4.4. Routeing/control dynamics under the equi-saturation control and the P0 control

In this section, we present numerical results showing the routeing/control dynamics under different control policies in the simple symmetric network of Fig. 2. We consider a symmetric network with K_1 = K2 = 1.1 (mins), $s_1 = s_2 = 30$ (veh/min) and $B = 1/2$. In this case, we assume that the total OD flow rate is $T < s = s_1 = s_2 = 30$ (veh/min).

We show in Appendix C a simple method to compute the trajectories of flow vectors arising from RPAP flow-swapping (2.1) and (2.2), and a responsive signal control policy.

Figure 3 plots the trajectories of flow vectors when the responsive equi-saturation policy and the delay formula (4.1) and (4.2) are followed. It shows that for any given demand (with equi-saturation) the set of consistent equilibria sometimes comprises three distinct points (one symmetrical equilibrium and two all or nothing equilibria) and sometimes comprises five distinct points (those mentioned above and also two further equilibria on the two "prongs" of the pitchfork). Moreover starting at a non-equilibrium a natural dynamical system will converge to one of the equilibria depending on the starting (flow, green-time) pattern; there is no guarantee of convergence to a single connected set of consistent equilibria as is the case with P_0. Further a small change in the starting position of the adjustment process may lead to convergence to a different consistent equilibrium; so a small change in the problem leads to a sharp change in the long run behaviour.

All these show how unpredictable the results of the routeing-control interactions are. Figure 3 also shows clearly the pitchfolk bifurcation, and confirms in great detail the suggestion in Smith and Mounce (2011); that even in the simplest signal-controlled network, stability is an issue when there is a responsive control system. The result demonstrates the instability and unpredictability arising with the equi-saturation policy.

Replacing the Webster's random delay function (4.2) with the first term of the P-K delay formula (4.4), we plot the trajectories of flow vectors when the equi-saturation policy is followed.

Figure 4 illustrates the effect of changing one of the parameters in the delay formula in the equi-saturation case; although the pitchfork has disappeared, the set of symmetrical equilibria is still unstable with trajectories diverging from it and converging to points where all flow is on just one of the two routes.

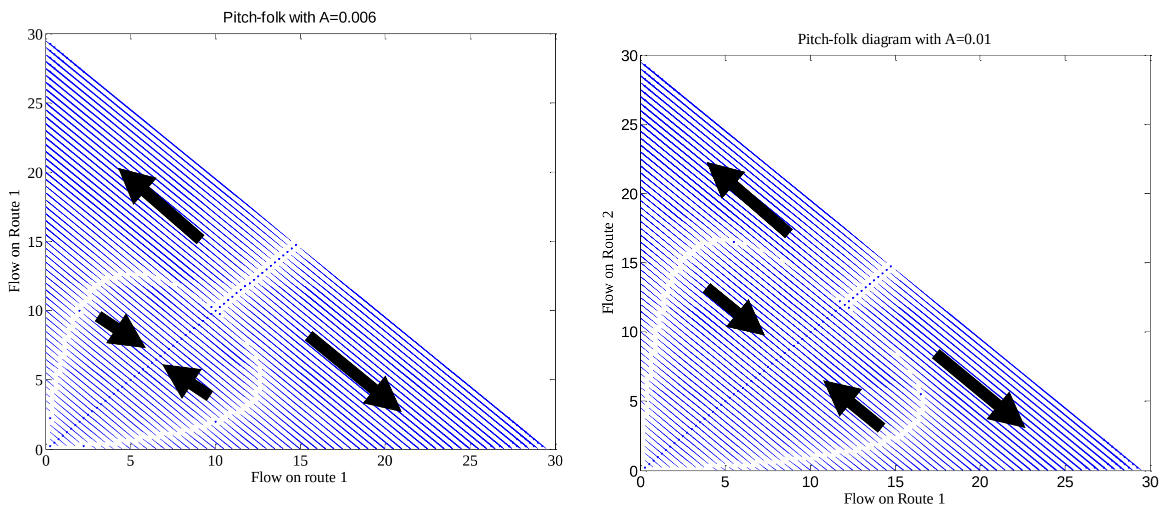

Figure 3. Using the equi-saturation policy with delay formula (4.1) and (4.2), a pitchfork-shaped set of equilibria arises. The dynamics of disequilibria are shown for two values of the slope A of the linear part of the cost flow function: (a) A=0.006 min/veh and (b) A=0.01 min/veh.

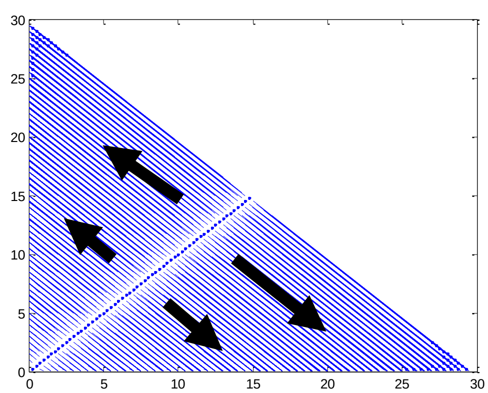

Figure 4. Using the equi-saturation policy with the delay formula (4.1) and (4.4), and A=0.006min/veh.

Figure 5 presents the trajectories of the flow vector when the P0 policy and a special delay function (which is a non-decreasing function of the red-time proportion) are followed. In contrast to Figures 3 and 4; Figure 5 illustrates that with P_0 all flow trajectories converge to the symmetrical set of equilibria. Thus in this special case the P_0 policy gives for each demand a single equilibrium state which is globally stable.

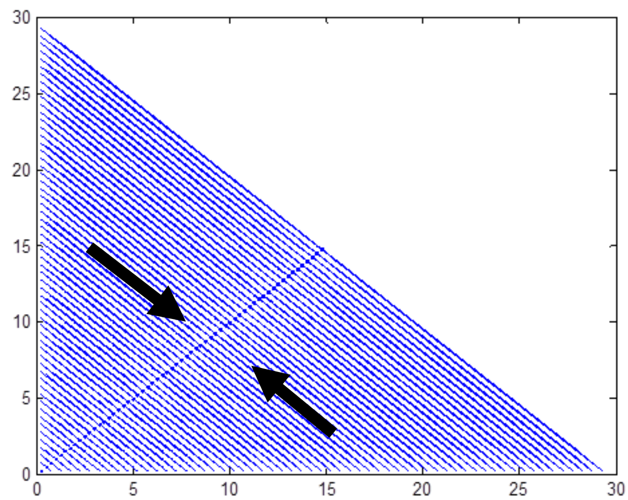

Figure 5. Using the P0 policy with the delay formula (4.1) and (4.4), and A=0.006 min/veh. B=1/2.

5. Conclusions

The paper has considered a new model of day-to-day re-routeing using restricted route-flow swaps following a restricted proportional-switch adjustment process (or RPAP). A corresponding dynamical model of green-times (or red-times) has been added. The central combined dynamical model in this paper is based on the special responsive control policy P_0 introduced in Smith (1979a, b) and cost functions, giving link costs in terms of flows and red-times, which are of a specific form.

Control dynamics in this paper have been stated in terms of the red-times allocated to "anti-stages" and links. This has allowed us to combine link red times and link traffic flows; leading to a two-commodity link model in which both the traffic flow and the red-time on a link contribute to the delay on that link. Having done this, similar proportional adjustment formulae have been utilised to specify both the routeing dynamics and the control dynamics. Both routeing and control adjustments are based on the proportional adjustment process (PAP) suggested in Smith (1984a) and discussed by He et al (2010); although here the PAP route-swap process has been modified (to RPAP) in light of the comments of He et al.

It has been shown that (under natural conditions) the discrete dynamical routeing model alone enters a set of approximate equilibria. This follows the continuous version in Smith (1984a). This central stability result is shown to hold also when the special dynamical P_0 green-time or red-time dynamical model is added to the dynamical routeing model. In this case the discrete routeing / control adjustment enters a set of approximate equilibria consistent with the P_0 policy.

The paper ends with examples showing that the equi-saturation policy may cause the joint (route flow, green time) dynamical system to be unstable and may give rise to the pitchfork bifurcation; this is done by plotting the route-swap directions induced by the equi-saturation policy. In these figures the pitchfork bifurcation appears clearly and it is obvious that the stability of the routeing / control dynamical system is, with equi-saturation, very unpredictable. Corresponding computational results with P_0 instead of equi-saturation demonstrate stability and predictability.

The paper suggests many questions which may be pursued. These include:

(a) How do the dynamical routeing systems studied here connect to other day to day dynamical routeing systems?

(b) Can a similar stability result be proved for other day to day dynamical systems,

(c) Can the P_0 policy be combined as here with other day to day dynamical systems ?

(d) If this is done, does the joint routeing / control dynamical system have a similar stability property as that demonstrated here with RPAP and P_0, perhaps utilising a similar two commodity cost function

(e) Can the signal control model in this paper be extended to include a wider class of realistic junction movements (such as lanes which allow left-turns and through movements), and more complex signal phasing schemes? The signal swapping model in this paper relies heavily on a degree of separability of the link-based red time; but in realistic junctions such separability is often missing. Applications to many real networks require more research in this direction.

In addition to extensions of this work in the directions suggested above, the route swap and red time swap models here may perhaps be extended (a) to within-day control/routeing systems and then (b) to allow time-varying demands. In order to do this it will be important to deal correctly with flow-propagation constraints and to adopt suitable travel time functions. This may well be difficult, but if successful would lead to large gains in applicability of the ideas in this paper.

Appendix A. Routeing and signal control: a modelling context

This appendix gives a short context concerning route choice modelling and traffic control modelling.

A.1 Route choice modelling.

Route choice modelling has typically been concerned with the problem of estimating the equilibrium distribution of traffic over a given network. This has been considered in a vast number of papers and books; and seeks iterative methods which ensure that route choices in model iterations converge to equilibrium; without seeking to design the iterations so as to necessarily represent a realistic within-day or day to day dynamical system. The following references concern steady state modelling and constitute a very small proportion of the literature. Bar-Gera and Boyce (2003; 2006), Cantarella (1997), Charnes and Cooper (1961), Dafermos (1980), Dial (1971, 2006), Evans (1976), Larsson and Patriksson (1992), Lv et al (2007), Maher (1998), Patriksson (1994), Payne and Thompson (1975), Sheffi (1985), Szeto and Lo (2006), Yang et al (1994), Yang and Huang (2004), Smith (1984b, 2009).

At the same time, within-day and day to day dynamics of traffic re-routing has been considered by Bie and Lo (2010), Cantarella and Cascetta (1995), Flötteröd and Liu (2014), Liu et al (2006), Smith (1984a) and others. These papers do not involve RPAP.

A.2 Signal control modelling.

Webster (1958) considered ways of determining signal timings for a single isolated intersection using a model of an isolated junction. As a result of his theoretical and simulation studies he suggested that the equi-saturation 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 equi-saturation policy aims to choose signal green times so that the saturation ratios on the two approaches are equal. Evers and Proost (2015) demonstrated the clear benefit of intersection regulation by traffic signal over priority rules.

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. 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 at each individual junction.

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 et al, 1982; this started as a responsive version of TRANSYT and follows equi-saturation 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 (Gartner, 1983). See Wood (1993) for a helpful discussion of the various systems. Heydecker (2004) outline motivations and new possible approaches to adaptive signal control.

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.

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.

A.3 Route choice and signal control modelling

The effects of changing signal timings on route-choices (and other decisions by users) are typically ignored by signal control designers. It was first pointed out by Allsop (1974) and Gartner (1976) 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 equi-saturation policy is to be employed then it is reasonable to seek green-times and flows where

(1) green-times satisfy the equi-saturation 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. Much of the existing theoretical work on re-routeing / control interactions has focused on one particular dynamical system: this is the standard method of trying to achieve a consistent (green-time, flow) pair (**G***, **X***). For example, for the equisaturation signal control policy 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 dynamical system 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).

Combining signal control and route choice within theoretical models has been considered by many others: for example, see Yang and Yagar (1994, 1995), Meneguzzer (1996) and the review by Meneguzzer (1997). 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. Maher et al (2013) considered the stochastic re-routing of drivers in response to a signal timing plan and applied a noisy optimisation method to find the globally optimal fixed-time signal plans that take into account of random errors in the objective function.

Hu and Mahmasami (1997), Mahmassami and Liu (1999), and Huang et al (2008) consider these dynamical issues within context of intelligent transportation systems; including information availability, utilizing microsimulation models. Clegg et al (2001) and Smith (2006) have considered the bi-level optimisation of prices and signals. Recently, Han et al (2014) developed a continuum approximation to the binary on-and-off signal controls, which provides a natural pathway for the combination of dynamic traffic assignment with signal optimization.

Appendix B. Proofs of Lemmas 1, 2, 3, 4 and 5; and of Theorems 1

APPENDIX B. 1

Proof of Lemma 1

Proof of part (a) of Lemma 1.

Consider an $\mathbf{X} \in \mathrm{D}$, let $\boldsymbol{\delta}$ be any feasible direction from $\mathbf{X}$ and consider the change in V as $\mathbf{X}$ changes to $\mathbf{X}+\mathrm{h}\boldsymbol{\delta}$ where $\mathrm{h}>0$. Then, from equation (2.10),

$$\mathrm{V}(\mathbf{X}+\mathrm{h}\boldsymbol{\delta})-\mathrm{V}(\mathbf{X}) = \sum_{\{(r,s);r\sim s\}} (\mathrm{X}_r + \mathrm{h}\delta_r)[\mathrm{C}_r(\mathbf{X}+\mathrm{h}\boldsymbol{\delta})-\mathrm{C}_s(\mathbf{X}+\mathrm{h}\boldsymbol{\delta})]_+^2 - \sum_{\{(r,s);r\sim s\}} \mathrm{X}_r[\mathrm{C}_r(\mathbf{X})-\mathrm{C}_s(\mathbf{X})]_+^2$$

$$= \sum_{\{(r,s);r\sim s\}} \{(\mathrm{X}_r + \mathrm{h}\delta_r)[\mathrm{C}_r(\mathbf{X}+\mathrm{h}\boldsymbol{\delta})-\mathrm{C}_s(\mathbf{X}+\mathrm{h}\boldsymbol{\delta})]_+^2 - \mathrm{X}_r[\mathrm{C}_r(\mathbf{X})-\mathrm{C}_s(\mathbf{X})]_+^2\}$$

$$= \sum_{\{(r,s);r\sim s\}} \mathrm{h}\delta_r[\mathrm{C}_r(\mathbf{X}+\mathrm{h}\boldsymbol{\delta})-\mathrm{C}_s(\mathbf{X}+\mathrm{h}\boldsymbol{\delta})]_+^2 +$$

$$\sum_{\{(r,s);r\sim s\}} \mathrm{X}_r\{[\mathrm{C}_r(\mathbf{X}+\mathrm{h}\boldsymbol{\delta})-\mathrm{C}_s(\mathbf{X}+\mathrm{h}\boldsymbol{\delta})]_+^2 - [\mathrm{C}_r(\mathbf{X})-\mathrm{C}_s(\mathbf{X})]_+^2\}$$

$$= \sum_{\{(r,s);r\sim s\}} \mathrm{h}\delta_r[\mathrm{C}_r(\mathbf{X}+\mathrm{h}\boldsymbol{\delta})-\mathrm{C}_s(\mathbf{X}+\mathrm{h}\boldsymbol{\delta})]_+^2 +$$

$$\sum_{\{(r,s);r\sim s\}} \mathrm{X}_r\{[\mathrm{C}_r(\mathbf{X}+\mathrm{h}\boldsymbol{\delta})-\mathrm{C}_s(\mathbf{X}+\mathrm{h}\boldsymbol{\delta})]_+ - [\mathrm{C}_r(\mathbf{X})-\mathrm{C}_s(\mathbf{X})]_+\}\{[\mathrm{C}_r(\mathbf{X}+\mathrm{h}\boldsymbol{\delta})-\mathrm{C}_s(\mathbf{X}+\mathrm{h}\boldsymbol{\delta})]_+ + [\mathrm{C}_r(\mathbf{X})-\mathrm{C}_s(\mathbf{X})]_+\}$$

Hence, for small h:

$$[\mathrm{V}(\mathbf{X}+\mathrm{h}\boldsymbol{\delta})-\mathrm{V}(\mathbf{X})]/\mathrm{h} = \sum_{\{(r,s);r\sim s\}} \delta_r[\mathrm{C}_r(\mathbf{X}+\mathrm{h}\boldsymbol{\delta})-\mathrm{C}_s(\mathbf{X}+\mathrm{h}\boldsymbol{\delta})]_+^2 +$$

$$\sum_{\{(r,s);r\sim s\}} (\mathrm{X}_r/\mathrm{h})\{[\mathrm{C}_r(\mathbf{X}+\mathrm{h}\boldsymbol{\delta})-\mathrm{C}_s(\mathbf{X}+\mathrm{h}\boldsymbol{\delta})]_+ - [\mathrm{C}_r(\mathbf{X})-\mathrm{C}_s(\mathbf{X})]_+\}\{[\mathrm{C}_r(\mathbf{X}+\mathrm{h}\boldsymbol{\delta})-\mathrm{C}_s(\mathbf{X}+\mathrm{h}\boldsymbol{\delta})]_+ + [\mathrm{C}_r(\mathbf{X})-\mathrm{C}_s(\mathbf{X})]_+\}$$

Now let $\mathrm{h}\to 0$. Then

$$[\mathrm{V}(\mathbf{X}+\mathrm{h}\boldsymbol{\delta})-\mathrm{V}(\mathbf{X})]/\mathrm{h} \to \sum_{\{(r,s);r\sim s\}} \delta_r[\mathrm{C}_r(\mathbf{X})-\mathrm{C}_s(\mathbf{X})]_+^2 + \sum_{\{(r,s);r\sim s\}} \mathrm{X}_r\{[\mathrm{C}'_r(\mathbf{X};\boldsymbol{\delta})-\mathrm{C}'_s(\mathbf{X};\boldsymbol{\delta})]_+\}\{2[\mathrm{C}_r(\mathbf{X})-\mathrm{C}_s(\mathbf{X})]_+\}$$

$$= \sum_{\{(r,s);r\sim s\}} \delta_r[\mathrm{C}_r(\mathbf{X})-\mathrm{C}_s(\mathbf{X})]_+^2 + \sum_{\{(r,s);r\sim s\}} [-\mathbf{C}'(\mathbf{X};\boldsymbol{\delta})\cdot\Delta_{rs}\}\{2\mathrm{X}_r[\mathrm{C}_r(\mathbf{X})-\mathrm{C}_s(\mathbf{X})]_+\}$$

$$= \sum_{\{(r,s);r\sim s\}} \delta_r[\mathrm{C}_r(\mathbf{X})-\mathrm{C}_s(\mathbf{X})]_+^2 + \sum_{\{(r,s);r\sim s\}} [-\mathbf{C}'(\mathbf{X};\boldsymbol{\delta})]\cdot 2\mathrm{X}_r[\mathrm{C}_r(\mathbf{X})-\mathrm{C}_s(\mathbf{X})]_+\Delta_{rs}$$

$$= \sum_{\{(r,s);r\sim s\}} \delta_r[\mathrm{C}_r(\mathbf{X})-\mathrm{C}_s(\mathbf{X})]_+^2 + [-\mathbf{C}'(\mathbf{X};\boldsymbol{\delta})]\cdot \sum_{\{(r,s);r\sim s\}} 2\mathrm{X}_r[\mathrm{C}_r(\mathbf{X})-\mathrm{C}_s(\mathbf{X})]_+\Delta_{rs}$$

$$= \sum_{\{(r,s);r\sim s\}} \delta_r[\mathrm{C}_r(\mathbf{X})-\mathrm{C}_s(\mathbf{X})]_+^2 - 2[\mathbf{C}'(\mathbf{X};\boldsymbol{\delta})]\cdot\mathbf{U}(\mathbf{X})/\mathrm{k}$$

It follows that V has a directional derivative $\mathrm{V}'(\mathbf{X};\boldsymbol{\delta})$ at $\mathbf{X}$ in direction $\boldsymbol{\delta}$ and that

$$\mathrm{V}'(\mathbf{X};\boldsymbol{\delta}) = \sum_{\{(r,s);r\sim s\}} \delta_r[\mathrm{C}_r(\mathbf{X})-\mathrm{C}_s(\mathbf{X})]_+^2 - 2[\mathbf{C}'(\mathbf{X};\boldsymbol{\delta})]\cdot\mathbf{U}(\mathbf{X})/\mathrm{k} \qquad \text{(B.1)}$$

This proves part (a) of lemma 1. □

Proof of part (b) of lemma 1.

By (B.1), and since $\mathbf{C}'(\mathbf{X};\boldsymbol{\delta})$ is a continuous function of $(\mathbf{X};\boldsymbol{\delta})$,

$$\mathbf{C}'(\mathbf{X};\mathbf{U}(\mathbf{X}))$$

is a continuous function of **X.** Also **C** is continuous and so

$$V'(\mathbf{X};\mathbf{U}(\mathbf{X})) = \sum_{\{(r,s);r\sim s\}} U_r(\mathbf{X})[C_r(\mathbf{X}) - C_s(\mathbf{X})]_+^2 - 2[\mathbf{C}'(\mathbf{X};\mathbf{U}(\mathbf{X}))]\cdot\mathbf{U}(\mathbf{X})/k \qquad \text{(B.2)}$$

is a continuous function of **X**. This proves part (b) of lemma 1.

Proof of part (c) of lemma 1.

By monotonicity of **C**

$$\mathbf{C}'(\mathbf{X};\mathbf{U}(\mathbf{X}))\cdot\mathbf{U}(\mathbf{X}) \geq 0.$$

and so (B.2) now yields:

$$V'(\mathbf{X};\mathbf{U}(\mathbf{X})) = \sum_{\{(r,s);r\sim s\}} \mathbf{U}(\mathbf{X})_r[C_r(\mathbf{X}) - C_s(\mathbf{X})]_+^2 - 2[\mathbf{C}'(\mathbf{X};\mathbf{U}(\mathbf{X}))]\cdot\mathbf{U}(\mathbf{X})/k$$
$$\leq \sum_{\{(r,s);r\sim s\}} \mathbf{U}(\mathbf{X})_r[C_r(\mathbf{X}) - C_s(\mathbf{X})]_+^2.$$

It may be shown (see the lemma in the appendix of Smith (1984b)) that, at all supply-feasible **X**,

$$\sum_{\{(r,s);r\sim s\}} \mathbf{U}(\mathbf{X})_r[C_r(\mathbf{X}) - C_s(\mathbf{X})]_+^2 \leq -k \sum_{\{(r,s);r\sim s\}} X_r[C_r(\mathbf{X}) - C_s(\mathbf{X})]_+^3.$$

Hence:

$$V'(\mathbf{X};\mathbf{U}(\mathbf{X})) \leq -\sum_{\{(r,s);r\sim s\}} X_r[C_r(\mathbf{X}) - C_s(\mathbf{X})]_+^3 < 0$$

away from equilibrium since

$$\sum_{\{(r,s);r\sim s\}} X_r[C_r(\mathbf{X}) - C_s(\mathbf{X})]_+^3 > 0$$

away from equilibrium. This completes the proof of part (c) of lemma 1.□

APPENDIX B. 2

Proof of Lemma 2.

Let $\mathbf{X}^{\mathbf{0}}$ be feasible (or $\mathbf{X}^{\mathbf{0}} \in \{\mathbf{Y};\ \mathbf{AY} < \mathbf{s}\} \cap D$) and let $\mathbf{X}^* \in \{\text{bdry}\{\mathbf{Y};\ \mathbf{AY} < \mathbf{s}\}\} \cap D$. Now let **X** start at $\mathbf{X}^{\mathbf{0}}$ and move steadily toward **X*** (by smooth route-flow swapping so that **X** remains in $\{\mathbf{Y};\ \mathbf{AY} < \mathbf{s}\} \cap D$ throughout until **X*** is reached). To be definite we may suppose here that

$$\mathbf{X}(t) = (1 - t)\mathbf{X}^{\mathbf{0}} + t\mathbf{X}^* \text{ for } 0 \leq t \leq 1.$$

Suppose that at **X*** (at t = 1) exactly m > 0 link exits become saturated simultaneously. Then the link flow through each of these m exit bottlenecks increases as $\mathbf{X} \to \mathbf{X}^*$; since each of these bottlenecks is unsaturated at $\mathbf{X}^0$, unsaturated at $\mathbf{X}(t)$ if $0 \leq t < 1$ and saturated at $\mathbf{X}^* = \mathbf{X}(1)$. Thus for each of these m bottlenecks the total flow along all routes passing through that bottleneck must rise as t increases to 1. Adding over the m bottlenecks, the total of all the route flows through one or more of these m bottlenecks must also then rise as t increases to 1. It follows that there is at least one origin-destination pair p such that the total of the route flows joining OD pair p and passing through at least one of the m bottlenecks increases as **X** moves toward **X*** or as t increases to 1.

Let us suppose then that route r is one such route; that is: route r joins OD pair p, passes through at least one of the m bottleneck links and has increasing flow as $\mathbf{X} \to \mathbf{X}^*$. As we are only swapping route flows in moving along the line joining $\mathbf{X}^0$ and $\mathbf{X}^*$, it follows that the total of the route-flows joining OD pair p which miss all of the m bottlenecks must decrease. So there must be at least one route (route s say) which joins OD pair p and misses all the m bottlenecks (and whose route-flow decreases). It now follows (since this route s misses all the m bottlenecks) that the cost $C_s(\mathbf{X})$ of travel along this route s is bounded above as $\mathbf{X} \to \mathbf{X}^*$**;** because route s passes through no saturated bottlenecks at **X***.

On the other hand route r passes through a bottleneck link and so that link cost, and hence $C_r(\mathbf{X})$ tends to infinity and also there is a constant a such that $X_r > a > 0$ as $\mathbf{X} \to \mathbf{X}^*$, since X_r increases as $\mathbf{X} \to \mathbf{X}^*$, (and $C_s(\mathbf{X})$ is bounded above as $\mathbf{X} \to \mathbf{X}^*$).

It follows that $X_r[C_r(\mathbf{X}) - C_s(\mathbf{X})]$ and $[C_r(\mathbf{X}) - C_s(\mathbf{X})]$ both tend to infinity as $\mathbf{X} \to \mathbf{X}^*$ and hence that $X_r[C_r(\mathbf{X}) - C_s(\mathbf{X})][C_r(\mathbf{X}) - C_s(\mathbf{X})] = X_r[C_r(\mathbf{X}) - C_s(\mathbf{X})]^2$ must also tend to infinity. This then implies that

$$V(\mathbf{X}) = \sum_{\{(r,s);r\sim s\}} X_r[C_r(\mathbf{X}) - C_s(\mathbf{X})]_+^2 \tag{2.10}$$

tends to infinity as $\mathbf{X} \to \mathbf{X}^*$ too.

We have shown here that $V(\mathbf{X})$ tends to infinity as $\mathbf{X} \to \mathbf{X}^* \in \{\text{bdry}\{\mathbf{Y};\, \mathbf{AY} < \mathbf{s}\}\} \cap D$. This completes the proof of lemma 2. □

APPENDIX B. 3

Proof of Lemma 3

Let $\mathbf{X}+\mathbf{H}$ and $\mathbf{X}$ be both feasible. Let $\mathbf{x} = \mathbf{AX}$ and $\mathbf{h} = \mathbf{AH}$. Then $\mathbf{x}$ and $\mathbf{x}+\mathbf{h}$ are both feasible and so:

$$[\mathbf{C}(\mathbf{X}+\mathbf{H})-\mathbf{C}(\mathbf{X})]\cdot\mathbf{H}=[\mathbf{A}^{\mathrm{T}}\mathbf{c}(\mathbf{AX}+\mathbf{AH})-\mathbf{A}^{\mathrm{T}}\mathbf{c}(\mathbf{AX})]^{\mathrm{T}}\mathbf{H}$$
$$=[\mathbf{c}(\mathbf{AX}+\mathbf{AH})-\mathbf{c}(\mathbf{AX})]^{\mathrm{T}}[\mathbf{AH}]=[\mathbf{c}(\mathbf{x}+\mathbf{h})-\mathbf{c}(\mathbf{x})]^{\mathrm{T}}[\mathbf{h}]\geq 0$$

since $\mathbf{c}=\mathbf{c}(\mathbf{x})$ is monotone. Therefore $\mathbf{C}=\mathbf{C}(\mathbf{X})$ is monotone. This completes the proof of lemma 3. □

APPENDIX B. 4

Proof of lemma 4.

Let $\mathbf{X}+\mathrm{h}\mathbf{H}$ and $\mathbf{X}$ be both feasible for small $\mathrm{h}>0$. Then as $\mathrm{h}\to 0+$,

$$[\mathbf{C}(\mathbf{X}+\mathrm{h}\mathbf{H})-\mathbf{C}(\mathbf{X})]/\mathrm{h}=[\mathbf{A}^{\mathrm{T}}\mathbf{c}(\mathbf{AX}+\mathbf{A}\mathrm{h}\mathbf{H})-\mathbf{A}^{\mathrm{T}}\mathbf{c}(\mathbf{AX})]/\mathrm{h}$$
$$=\mathbf{A}^{\mathrm{T}}[\mathbf{c}(\mathbf{AX}+\mathbf{A}\mathrm{h}\mathbf{H})-\mathbf{c}(\mathbf{AX})]/\mathrm{h}\to\mathbf{A}^{\mathrm{T}}[\mathbf{c}'(\mathbf{AX},\mathbf{AH})]$$

since $\mathbf{c}$ is directionally differentiable. Therefore $\mathbf{C}=\mathbf{C}(\mathbf{X})$ is directionally differentiable at $\mathbf{X}$ in direction $\mathbf{H}$.

This directional derivative is $\mathbf{A}^{\mathrm{T}}[\mathbf{c}'(\mathbf{AX},\mathbf{AH})]$; and this is a continuous function of ($\mathbf{X}$, $\mathbf{H}$) as the directional derivative of **c** in direction **AH** at **AX** is a continuous function of (**AH**, **AX**). This completes the proof of lemma 4. □

APPENDIX B. 5

Proof of Theorem 1.

This uses Lemma 1 and a standard Lyapunov approach; similar to that utilised in the appendix in Smith and Mounce (2011).

Let all the conditions in the statement of theorem 1 hold, let $\mathbf{X}^0$ be feasible, let k satisfy the extended non-negativity result, let **U(X)** be then given by equation (2.8), and let $\varepsilon>0$. Then **C** is monotone on S ∩ D since the link cost function **c** is monotone on [**0**, **s**); therefore **U**(**X**) (given in (2.8)) is a descent direction for V at each $\mathbf{X}\in \mathrm{S}\cap\mathrm{D}$ such that V(**X**) > 0. Proof of (i). Here we show that continuity of **U**, C**'** and hence *V'* now yields the existence of $\mathrm{h}_0 = \mathrm{h}_0(\mathbf{X}^0)$ such that

$$\mathbf{X}\in \mathrm{D}^{00}\Rightarrow \mathbf{X}+\mathrm{h}\mathbf{U}(\mathbf{X})\in \mathrm{D}^{00} \text{ for all h such that } 0<\mathrm{h}<\mathrm{h}_0.$$

This will then have shown that (i) holds. Note that if $\mathbf{X}\in \mathrm{D}^{00}$ then

$$\mathbf{X}\in \mathrm{D}^{00}\cap\{\mathbf{X}\in \mathrm{S}\cap \mathrm{D};\ \mathrm{V}(\mathbf{X})\geq\varepsilon\} \text{ or } \mathbf{X}\in \mathrm{D}^{00}\cap\{\mathbf{X}\in \mathrm{S}\cap \mathrm{D};\ \mathrm{V}(\mathbf{X})\leq\varepsilon\}$$

and we consider these last two possibilities separately.

So first let $\mathbf{X} \in D^{00} \cap \{\mathbf{X} \in S \cap D;\ V(\mathbf{X}) \geq \varepsilon\}$. Then by lemma 1 (using monotonicity of **C**) inequality (2.12) holds. Since $D^{00} \cap \{\mathbf{X} \in S \cap D;\ V(\mathbf{X}) \geq \varepsilon\}$ is closed and bounded (and so compact), and since V'(.) is continuous on $D^{00} \cap \{\mathbf{X} \in S \cap D;\ V(\mathbf{X}) \geq \varepsilon\}$ convergence in (2.12) is uniform on $D^{00} \cap \{\mathbf{X} \in S \cap D;\ V(\mathbf{X}) \geq \varepsilon\}$. So there exists $h_1 \leq 1$ such that

$$0 < h \leq h_1 \text{ implies } [V(\mathbf{X} + h\mathbf{U}(\mathbf{X})) - V(\mathbf{X})] / h < V'(\mathbf{X}, \mathbf{U}(\mathbf{X}))/2 < 0$$

for all $\mathbf{X} \in D^{00} \cap \{\mathbf{X} \in S \cap D;\ V(\mathbf{X}) \geq \varepsilon\}$. Here h_1 does not depend on $\mathbf{X} \in D^{00} \cap \{\mathbf{X} \in S \cap D;\ V(\mathbf{X}) \geq \varepsilon\}$. It follows that, for all $\mathbf{X} \in D^{00} \cap \{\mathbf{X} \in S \cap D;\ V(\mathbf{X}) \geq \varepsilon\}$,

$$0 < h \leq h_1 \text{ implies } V(\mathbf{X} + h\mathbf{U}(\mathbf{X})) < V(\mathbf{X}) < V(\mathbf{X}^0)$$

and so $\mathbf{X} + h\mathbf{U}(\mathbf{X})) \in D^{00} \cap (S \cap D)$.

Now, on the other hand, let $\mathbf{X} \in D^{00} \cap \{\mathbf{X} \in S \cap D;\ V(\mathbf{X}) \leq \varepsilon < V(\mathbf{X}^0)\}$. Then, since V and **U** are both continuous and $V(\mathbf{X}^0) - \varepsilon > 0$, there exists h_2 such that

$$0 < h \leq h_2 \text{ implies } V(\mathbf{X} + h\mathbf{U}(\mathbf{X})) - V(\mathbf{X}) < V(\mathbf{X}^0) - \varepsilon$$

which implies that $V(\mathbf{X} + h\mathbf{U}(\mathbf{X})) < V(\mathbf{X}^0) - \varepsilon + V(\mathbf{X}) \leq V(\mathbf{X}^0)$ and hence that

$$\mathbf{X} + h\mathbf{U}(\mathbf{X})) \in D^{00} \cap (S \cap D).$$

To combine these two results let $0 < h \leq h_0 = \min\{h_1, h_2\}$. Then both of the two implications above hold and so

$$\mathbf{X} \in D^{00} \cap (S \cap D) \text{ implies } \mathbf{X} + h\mathbf{U}(\mathbf{X}) \in D^{00} \cap (S \cap D)$$

and this is (i).

Proof of (ii). Now let h_0 satisfy (i). Then the infinite sequence (2.15) is properly defined for any h such that $0 < h \leq h_0$. The approximate equilibrium set $E_\varepsilon = \{\mathbf{X} \in S \cap D;\ V(\mathbf{X}) \leq \varepsilon\}$ and we need now to check that if $0 < h \leq h_0$ the sequence (2.15) also certainly enters E_ε.

Let $\mathbf{X} \in D^{00} \cap \{\mathbf{X} \in S \cap D;\ V(\mathbf{X}) \geq \varepsilon\}$. Then by the argument above

$$0 < h \leq h_0 \text{ implies } [V(\mathbf{X} + h\mathbf{U}(\mathbf{X})) - V(\mathbf{X})] / h < V'(\mathbf{X}, \mathbf{U}(\mathbf{X}))/2 < 0$$

for all $\mathbf{X} \in D^{00} \cap \{\mathbf{X} \in S \cap D;\ V(\mathbf{X}) \geq \varepsilon\}$. Now, since $D^{00} \cap \{\mathbf{X} \in S \cap D;\ V(\mathbf{X}) \geq \varepsilon\}$ is closed and bounded (and so compact) and V' and **U** are continuous, there is $q > 0$ such that

$$V'(\mathbf{X}, \mathbf{U}(\mathbf{X}))/2 \leq -q < 0$$

for all $\mathbf{X}$ in $D^{00} \cap \{\mathbf{X} \in S \cap D;\ V(\mathbf{X}) \geq \varepsilon\}$. Thus if h is fixed, $0 < h \leq h_0$ and

$$T(\mathbf{X}) = \mathbf{X} + h\mathbf{U}(\mathbf{X}),$$

for all $\mathbf{X}$ in $D^{00} \cap \{\mathbf{X} \in S \cap D;\ V(\mathbf{X}) \geq \varepsilon\}$, then:

$$V(T(\mathbf{X})) - V(\mathbf{X}) = [V(\mathbf{X} + h\mathbf{U}(\mathbf{X})) - V(\mathbf{X})] < hV'(\mathbf{X}, \mathbf{U}(\mathbf{X}))/2 \le - hq < 0$$

for all $\mathbf{X}$ in $D^{00} \cap \{\mathbf{X} \in S \cap D;\ V(\mathbf{X}) \ge \varepsilon\}$. It follows that if

$$\mathbf{X}^0, T\,\mathbf{X}^0, T^2\,\mathbf{X}^0, \ldots, T^{n-1}\,\mathbf{X}^0, T^n\,\mathbf{X}^0$$

all belong to $D^{00} \cap \{\mathbf{X} \in S \cap D;\ V(\mathbf{X}) \ge \varepsilon\}$ then:

$$\begin{aligned}
&V(T^n\,\mathbf{X}^0) - V(\mathbf{X}^0)\\
&= [V(T^n\,\mathbf{X}^0)-V(T^{n-1}\,\mathbf{X}^0)] + [[V(T^{n-1}\,\mathbf{X}^0)-V(T^{n-2}\,\mathbf{X}^0)] +\ldots+ [V(T^2\,\mathbf{X}^0)-V(T\,\mathbf{X}^0)] + [V(T\,\mathbf{X}^0)-V(\mathbf{X}^0)]\\
&< \quad - hq \qquad\qquad - hq \qquad\qquad - hq \qquad\qquad - hq\\
&= - nhq.
\end{aligned}$$

It now follows that in this case:

$$V(T^n\,\mathbf{X}^0) < V(\mathbf{X}^0) - nhq < \varepsilon$$

if $n > [V(\mathbf{X}^0) - \varepsilon] / hq$. Hence, if $n > [V(\mathbf{X}^0) - \varepsilon] / hq$,

$$\mathbf{X}^0, T\,\mathbf{X}^0, T^2\mathbf{X}^0, \ldots, T^{n-2}\,\mathbf{X}^0, T^{n-1}\mathbf{X}^0, T^n\mathbf{X}^0$$

cannot all belong to $D^{00} \cap \{\mathbf{X} \in S \cap D;\ V(\mathbf{X}) \ge \varepsilon\}$.

Hence the infinite sequence (2.15) certainly enters E_ε: the above inequality shows that $T^n\mathbf{X}^0$ must first enter E_ε when n first exceeds $[V(\mathbf{X}^0) - \varepsilon] / hq$, at the very latest.

Theorem 1 is proved.□

APPENDIX B. 6

Proof of lemma 5.

Suppose that c_i and b_i are monotone cost functions associated with link i; here we drop the suffices. Suppose that x, $x+\delta x$, $x+sr$ and $x+\delta x+sr+s\delta r$ are all feasible. Then

$$\begin{aligned}
&\{[c(x+\delta x)+b(x+\delta x+sr+s\delta r), sb(x+\delta x)+s(r+\delta r)]-[c(x)+b(x+sr), sb(x+sr)]\}\cdot[\delta x, \delta r]\\
&=[c(x+\delta x)-c(x)]\cdot\delta x+[b(x+\delta x+sr+s\delta r)-b(x+sr)]\cdot\delta x+[sb(x+\delta x)+s(r+\delta r)-sb(x+sr)]\cdot\delta r\\
&=[c(x+\delta x)-c(x)]\cdot\delta x+[b(x+sr+(\delta x+s\delta r))-b(x+sr)]\cdot[\delta x+\delta r]\\
&\ge \quad 0 \qquad\qquad + \qquad\qquad 0\\
&\ge 0,
\end{aligned}$$

since c and b are both monotone. This completes the proof of lemma 5. □

Appendix C. Calculation of the trajectories of flow vector following RPAP route-flow swap and a responsive signal control policy

A simple method to examine the stability of a dynamical system of route flow swapping and responsive signal control is to follow the trajectories of flows from neighbouring supply-feasible regions and to see how the dynamical system behave.

For a given network, e.g. the example network of Figure 2, starting from a feasible flow vector $(\mathrm{T}, \mathbf{H})$, we follow an iterative process of route-flow swapping and signal green (or red) proportion adjustment, until no more changes can be made to either the route flows or the green/red signal proportions. The details of the calculation method is outlined as follows:

Step 0 **Input**

Input: Network description and link cost function variables: s_1, s_2, K_1, K_2, A, B

Input: step size k, and a flow-swap tolerance level ε

Select: responsible control policy and the delay formula

For each T=1, 2, 3, …. max(s_1, s_2)

Step 1 **Initialisation**

Set day counter t=1. Choose initial flow vector $(\mathrm{T}, \mathbf{H})$ that satisfies the flow feasibility condition (4.8)

Step 2 **Compute the route flow and route costs**

Calculate route flow $\mathbf{X} = \mathrm{T}\mathbf{H}$

Calculate route cost $\mathbf{C}$ from (4.1) choosing the steep delay formula following either the Webster's random delay term (4.2) or the first term of P-K (4.4)

Step 3 **Route flow swap**

Compute the amount of route flow swap $\mathbf{U}$ according to (2.1) and (2.2)

Step 4 **Solution improvement check**

If the amount of flow swap is less than the predefined value ε, then stop and report the final route flow X and green/red proportions G/R

Otherwise, compute the new flow $\mathbf{X}(t+1) = \mathbf{X}(t) + \mathbf{U}(\mathbf{X}(t))$ and new flow split $\mathrm{H}_i(t+1) = \mathrm{X}_i(t+1) / \sum \mathrm{X}_j(t+1)$, set t= t+1, go to Step 2.

References

Allsop, R. E., 1974. Some possibilities for using Traffic Control to Influence Trip Distribution and Route Choice. Proceedings of the 6th International Symposium on Transportation and Traffic Theory, Sydney (Editor: Buckley, D. J.), Elsevier, 345-374.

Bar-Gera, H., Boyce, D., 2003. Origin-based algorithms for combined travel forecasting models. Transportation Research Part B, 37 (5), 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 (5), 351-367.

Bar-Gera, H., 2010. Traffic assignment by paired alternative segments. Transportation Research Part B, 44 (8-9), 1022-1046.

Bellei, G., Gentile, G. and Papola, N., 2005. A within-day dynamic traffic assignment model for urban road networks. Transportation Research Part B, 39 (1), 1–29

Bentley, R. W., Lambe, T. A., 1980. Assignment of traffic to a network of signalized city streets. Transportation Research Part A, 14 (1), 57-65.

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(1), 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., E. 2010. Signal Setting with Dynamic Process Assignment. New developments in transport planning, (Eds: Tampere, M. J., Viti, F., Immers, L. H.), Edward Elgar, 29 – 56.

Cantarella, G. E., Cascetta, E. 1995. Dynamic processes and equilibrium in transportation networks: towrds a unifying theory. Transportation Science 29(4), 305-329.

Cantarella, G. E., Improta, G., Sforza, A. 1991. Road network signal setting: equilibrium conditions. Concise Encyclopedia of Traffic and Transportation Systems (Editor: Papageorgiou, M), Pergamon, 366 – 371.

Cascetta, E. 1989. A stochastic process approach to the analysis of temporal dynamics in transportation networks, Transportation Research Part B, 23(1), 1-17.

Charnes, A., Cooper, W. W., 1961. Multicopy traffic network models. Theory of Traffic Flow. (Proceedings of the Symposium on the Theory of Traffic Flow, General Motors Research Laboratories, Warren, MI, 1959. Editor: R. Herman.) Elsevier, 85-96.

Clegg, J., Smith, M. J., Xiang, Y., Yarrow, R., 2001. Bilevel programming applied to optimising urban transportation. Transportation Research Part B, 35 (1), 41-70.

Dafermos, S., 1980. Traffic equilibrium and variational inequalities. Transportation Science 14 (1), 42 – 54.

Dial, R., 1971. A probabilistic multipath assignment model that obviates path enumeration. Transportation Research 5 (2), 83-111.

Dial, R., 2006. A path-based user-equilibrium assignment algorithm that obviates path storage and enumeration. Transportation Research Part B 40 (10), 917 – 936.

Dickson, T. J., 1981. A note on traffic assignment and signal timings in a signal-controlled road network. Transportation Research Part B 15 (4), 267 – 271.

Evans, S. P., 1976. Derivation and analysis of some models for combining trip distribution and assignment. Transportation Research 10 (1), 37-57.

Evers, R., Proost. S., 2015. Optimizing intersections. Transportation Research B71, 100-119.

Flötteröd, G. and Liu, R (2014) Disaggregate path flow estimation in an iterated DTA microsimulation. Journal of Intelligent Transportation Systems: Technology, Planning and Operations, 18(2), 204-214.

Gartner, N. H. 1976. Area Traffic Control and Network Equilibrium, Traffic Equilibrium Methods, Lecture Notes in Economics and Mathematical Systems., 118, (Editor: Florian, M.) Springer Verlag, 274 – 297.

Gartner, N. H., 1983. OPAC: A demand responsive strategy for traffic signal control. Transportation Research Record 906, 75 - 81.

Han, K., Gayah, V.V., Piccoli, B., Friesz, T.L., Yao, T., 2014. On the continuum approximation of the on-and-off-signal control on dynamic traffic networks. Transportation Research, B 61, 73-97.

He, X., Guo, X., Liu, H., 2010. A link-based day to day traffic assignment model. Transportation Research Part B, 44 (4), 597 – 608.

Heydecker, B. G., 2004. Objectives, stimulus and feedback in signal control of road traffic. ITS Journal 8 (2), 63-76

Hu, T-Y., Mahmassami, H. S., 1997. Day to day evolution of network flows under real-time information and reactive signal control. Transportation Research C5(1), 51 – 69.

Huang, H.J., Lam, W., 2002. Modeling and solving the dynamic user equilibrium route and departure time choice problem in network with queues. Transportation Research B36 (3), 253 – 273.

Huang, H. J., Liu, T.L., Yang, H. 2008. Modeling the evolutions of day-to-day route choice and year-to-year ATIS adoption with stochastic user equilibrium. Journal of Advanced Transportation, Vol 42(2), 111 - 127.

Hunt, P.B., Robertson, D.I., Bretherton, R.D., Boyle, M.C., 1982. The SCOOT on-line traffic signal optimisation technique. Traffic Engineering and Control 23, 190-192.

Khintchine, A. Y (1932). "Mathematical theory of a stationary queue". Matematicheskii Sbornik 39 (4): 73–84. Retrieved 2011-07-14.

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).

Liu, R., van Vliet, D., Watling, D. 2006. Microsimulation models incorporating both demand and supply dynamics. Transportation Research, Part A, 40, 125-150.

Lv, Q., Salis, A., Skrobanski, G., Smith, M.J., Springham, J., Woods, A., Gordon, A., 2007. Two-direction methods for variable demand traffic assignment. Selected Proceedings of 4th IMA International Conference on Mathematics in Transport. (Ed: B. Heydecker), Elsevier, 71-95.

Lyapunov, A. M., 1907. Probleme general de la stabilite de mouvement, Ann. Fac. Sci. Toulouse 9, 203 – 474. Reprinted in Ann. Math. Stud. No 17 (Princeton University Press), 1949.

Madan, K.C. and Saleh, M.H., 2001. On M/D/1 queue with geneal server vacations. Information and Management Sciences. Vol 12(2), 25-37.

Mahmassami, H. S., Liu, Y-H, 1999. Dynamics of commuting decision behaviour under advanced traveler information systems. Transportation Research Part C, 7 (2-3), 91-107.

Maher, M. J., 1998. Algorithms for logit-based stochastic user equilibrium assignment. Transportation Research Part B 32 (8), 539 – 549.

Maher, M. J., Zhang, X., van Vliet, D. 2001. A Bi-level Programming approach for Trip Matrix Estimation and Traffic Control Problems with Stochastic User Equilibrium Link Flows. Transportation Research Part B, 35 (1), 23-40.

Maher, M., Liu, R. and Ngoduy, D. (2013) Signal optimisation using the cross entropy method. Transportation Research Part C, Vol 27, 76-88.

Meneguzzer, C., 1996. Computational experiments with a combined traffic assignment and control model with asymmetric cost functions. Proceedings of the 4th International Conference on Applications of Advanced Technologies in Transportation Engineering (Editors: Stephanedes, Y. J., Filippi, F.), ASCE, New York, 609-614.

Meneguzzer, C., 1997. Review of models combining traffic assignment and signal control. Journal of Transportation Engineering, 123, 148-155.

Mounce, R., 2006. Convergence in a continuous dynamic queueing model for traffic networks. Transportation Research Part B, 40, 779–791.

Mounce, R., 2009. Existence of equilibrium in a continuous dynamic queueing model for traffic networks with responsive signal control. Proceedings of 18th International Symposium on Transportation and Traffic Theory, (Eds: Lam, WHK, Wong, SC, Lo, HK), Springer, 327-344.

Mounce, R. and Carey, M., 2011. Route swapping in dynamic traffic networks. Transportation Research Part B 45 (1) 102–111.

Mounce, R. and Smith, M., 2007. Uniqueness of equilibrium in steady state and dynamic traffic networks. In: Transportation and Traffic Theory, 281–299, Elsevier, Eds: R.E.Allsop, M.G.H.Bell and B.G.Heydecker.

Nie, Y., 2010. A class of bush-based algorithms for the traffic assignment problem. Transportation Research Part B 44 (1), 73 – 89.

Nie, X. and Zhang, H. M. (2005). A comparative study of some macroscopic link models used in dynamic traffic assignment. Networks and Spatial Economics 5(1), 89-115.

Patriksson, M., 1994. The traffic assignment problem – Models and Methods. VSP, Utrecht, Netherlands.

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.

Peeta, S. Yang, T.H. 2003. Stability issues for dynamic traffic assignment. Automatica 39(1), 21 - 34.

Pollaczek, F. (1930). "Über eine Aufgabe der Wahrscheinlichkeitstheorie". Mathematische Zeitschrift 32: 64–100. doi:10.1007/BF01194620.

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., 1979a. 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, Volume 2A, 11-32; and in Traffic Engineering and Control, 21, 1980, 298 - 302.

Smith, M. J. 1979b. The existence, uniqueness and stability of traffic equilibria. Transportation Research 13B, 295-304.

Smith, M. J. 1979c. Traffic control and route-choice; a simple example. Transportation Research 13B, 289-294.

Smith, M. J., 1984a. The stability of a dynamic model of traffic assignment - an application of a method of Lyapunov. Transportation Science 18 (3), 245 - 252.

Smith, M. J., 1984b. A descent algorithm for solving a variety of monotone equilibrium problems. Proceedings of 9th International Symposium on Transportation and Traffic Theory, Delft, Netherlands (Eds: Volmuller, J., Hamerslag, R.), VNU Science Press, 273-297.Smith, M. J., Van Vuren, T., Van Vliet, D and Heydecker, B. G. (1987). The interaction between signal control policies and route choice, Proceedings of the Tenth International Symposium on Transportation and Traffic Theory (Editors: Nathan H. Gartner and Nigel H. M. Wilson), Elsevier, 319-338.

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 (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 (2), 118-132.

Smith, M. J., Wisten, M. B., 1995, A continuous day-to-day traffic assignment model and the existence of a continuous dynamic user equilibrium. Annals of Operations Research 60, 59-79.

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 18th International Symposium on Transportation and Traffic Theory, Hong Kong Polytechnic University, Hong Kong. (Eds: Lam, W. H. K., Wong, S. C., Lo, H. K.), Springer, 365 - 386.

Smith, M. J. , Mounce, R. (2011) A splitting rate model of traffic re-routeing and traffic control. Transportation Research Part B, 1389 - 1409.

Szeto, W.Y., Lo, H., 2006. Dynamic traffic assignment: properties and extensions. Transportmetrica, 2, 31 - 52.

Taale, H., van Zuylen, H.J. 2001. The combined traffic assignment and control problem – an overview of 25 years of research. Proceedings of the World Conference on Transport Research, Seoul, 22 - 27.

Wardrop, J. G., 1952. Some Theoretical Aspects of Road Traffic Research, Proceedings, Institution of Civil Engineers II (1), 235-278.

Webster, F. V., 1958. Traffic Signal Settings, Department of Transport, Road Research Technical Paper No. 39, HMSO, London.

Wong, S.C., Yang, C., Lo, H., 2001. A path-based traffic assignment algorithm using the TRANSYT traffic model. Transportation Research Part B 35, 163 - 181.

Wood, K., 1993. Urban Traffic Control, Systems Review. Project Report 41, Transport Research Laboratory, Crowthorne, United Kingdom.

Yang, H., Huang, H. J., 2004. Modeling user adoption of advanced traveler information systems: A control theoretical approach for optimal endogenous growth. Transportation Research Part C 12 (3-4), 193 - 207.

Yang, H., Yagar, S., 1994. Traffic assignment and traffic control in general freeway-arterial corridor systems. Transportation Research Part B 28 (6), 161 - 174.

Yang, H., Yagar, S., 1995. Traffic assignment and signal control in saturated road networks. Transportation Research Part A 29 (2), 125 - 139.

Yang, H., Yagar, S., Iida, Y. 1994. Traffic assignment in a congested discrete-continuous transportation system. Transportation Research Part B 28 (2), 161 - 174.