

THE STABILITY OF DOWNTOWN PARKING AND TRAFFIC CONGESTION*

Richard Arnott[†]

University of California, Riverside

Eren Inci[‡]

Sabanci University

27 November 2008

Abstract

In classical traffic flow theory, there are two velocities associated with a given level of traffic flow. Following Vickrey, economists have termed travel at the higher speed congested travel and at the lower speed hypercongested travel. Since the publication of Walters' classic paper (1961, *Econometrica* **29**, 676-699), there has been an on-going debate concerning whether a steady-state hypercongested equilibrium can be stable. For a particular structural model of downtown traffic flow and parking, this paper demonstrates that a steady-state hypercongested equilibrium can be stable. Some other sensible models of traffic congestion conclude that steady-state hypercongested travel cannot be stable, and that queues develop to ration the demand in steady states. Thus, we interpret our result to imply that, when steady-state demand is so high that it cannot be rationed through congested travel, the trip price increase

*The authors would like to thank Albert Erkip, Thomas Holmes, Robin Lindsey, Kenneth Small, Mete Soner, Erik Verhoef, and the participants at the 3rd International Conference on Funding Transport Infrastructure and 10th Journée Transport, the Macroeconomics, Real Estate, and Public Policy Workshop, and the 55th Annual North American Meetings of the Regional Science Association International, especially the discussant Jeffrey Lin, for valuable comments.

[†]*Address:* Department of Economics, University of California, Riverside, 4106 Sproul Hall, Riverside, CA 92521-0427 USA. *E-mail address:* richard.arnott@ucr.edu

[‡]*Corresponding Author. Tel.:* 90-216-483-9340; *fax:* 90-216-483-9250. *Address:* Sabanci University - FASS, Orhanli / Tuzla 34956 Istanbul TURKEY. *E-mail address:* ereninci@sabanciuniv.edu.

necessary to ration the demand may be generated either through the formation of steady-state queues or through hypercongested travel, and that which mechanism occurs depends on details of the traffic system.

Keywords: traffic congestion, cruising for parking, on-street parking, hypercongestion

JEL Classification: R41, L91

1 Introduction

To non-experts, many academic debates seem arcane. The amount of ink spent on them seems quite out of proportion to the importance of the issues under debate. But more often than not, the debates provide a focal point for discussion about the fundamentals of a field. In transport economics, there has been only one major theoretical debate, which has been active for almost fifty years. In classical traffic flow theory (known also as kinematic wave theory or Lighthill-Whitham-Richards (LHR) traffic flow theory), there are two velocities associated with a given level of flow (for example, zero flow corresponds to no cars on the road and a complete traffic jam). Following Vickrey, economists have termed travel at the higher velocity congested travel and at the lower velocity hypercongested travel. With hypercongested travel, an increase in flow is associated with an *increase* in velocity – the unjamming of a traffic jam. Since the seminal article by Walters (1961), the transport economics literature has debated whether there exist steady-state equilibria with the comparative static property that, in response to an exogenous change in demand, the change in flow is positively related to the change in velocity¹, and if such equilibria exist whether they are stable. Put informally, can steady-state traffic behave like a traffic jam? The issue is fundamental since it concerns the modeling of traffic congestion, which is central to transport economic theory.

In their recent magisterial textbook, two of the most distinguished transport economic theorists, Kenneth Small and Erik Verhoef (2007, ps. 84-86) have proposed a resolution of the debate, which builds on a series of papers (Verhoef, 1999, 2001,

¹Significant contributions to the earlier debate, listed according to date of publication, are Johnson (1964), Neuberger (1971), Agnew (1977), Dewees (1978), Else (1981,1982), Nash (1982), and McDonald and d’Ouille (1988)

and 2003; and Small and Chu, 2003) written over several years, during the course of which the authors' thinking on the subject evolved. Their proposed resolution, which we shall explain in greater detail in the next section, is that the backward-bending portion of the steady-state user cost curve should be replaced by a vertical section, corresponding to a 'vertical' queue. Consider a road system in which steady-state demand is so high that it cannot be rationed through congested travel. They argue that trip price increases to ration the demand through the formation of steady-state queues rather than through hypercongested travel.²

In this paper, we examine the stability of steady-state equilibria in a structural model of downtown parking and traffic congestion (Arnott and Inci, 2006) through a detailed analysis of the model's non-stationary dynamics. We show that, *in the context of this model and with the type of stability we consider*, there do exist stable, hypercongested, steady-state equilibria. Since our model is not general, we interpret our result to imply that, when steady-state demand is so high that it cannot be rationed through congested travel, the trip price increase necessary to ration the demand may be generated either through the formation of steady-state queues or through hypercongested travel, and that which mechanism occurs depends on details of the traffic system.

Section 2 provides a technical statement of the debate over the existence of a stable, hypercongested, steady-state traffic equilibrium, and reviews the relevant literature. Section 3 describes the model. Section 4 derives the model's steady-state equilibria and identifies which are hypercongested. Section 5 investigates the stability of the equilibria. Section 6 discusses the results.

2 Do There Exist Stable, Hypercongested, Steady-State Traffic Equilibria? A Review of the Debate

Imagine a homogeneous road between two locations with a constant flow of cars along it. Denote flow with f . Assume that traffic congestion is described by a technological

²They develop their argument for a straight highway but we at least interpret them as implying that their proposed resolution applies to traffic systems generally.

relationship between velocity, v , and density, V , with velocity being inversely related to density. For the sake of concreteness, assume Greenshield's Relation (1935), which specifies that there is a negative linear relationship between velocity and density:

$$v = v_f(1 - \frac{V}{V_j}) \quad \text{or} \quad V = V_j(1 - \frac{v}{v_f}) \quad , \quad (i)$$

where v_f is free-flow velocity and V_j is jam density. The Fundamental Identity of Traffic Flow is that flow equals velocity times density:

$$f = V \times v \quad . \quad (ii)$$

Combining (i) and (ii) gives flow as a function of velocity:

$$f = \frac{V_j(v_f - v)v}{v_f} \quad , \quad (iii)$$

which is an inverted and translated parabola and is displayed in Figure 1.

Maximum flow is referred to as capacity (flow). There are two velocities associated with each level of flow below capacity flow. Following Vickrey, economists refer to travel at the higher velocity as *congested* traffic flow and travel at the lower velocity as *hypercongested* flow. Congested traffic flow is informally interpreted as smooth-flowing traffic and hypercongested traffic flow as a traffic jam situation.

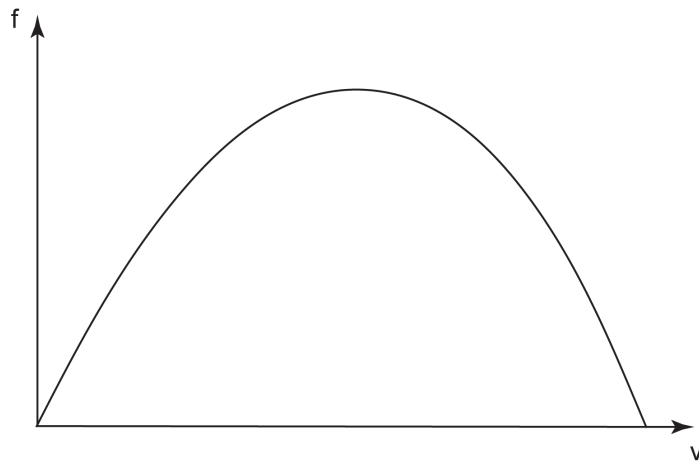


Figure 1: Flow as a function of velocity

Assume to simplify that the money costs of travel are zero and that the value of travel time is independent of traffic conditions and is the same for all cars. Then the cost of a trip, c , is simply the value of travel time, ρ , times travel time, t , which is the inverse of speed, times the length of the street, which we normalize to one, without loss of generality: $c = \rho t = \rho/v$ or $v = \rho/c$. Substituting this into (iii) gives the relationship between trip cost and flow:

$$f = \frac{V_j(v_f c - \rho)\rho}{v_f c^2} . \quad (iv)$$

Figure 2 plots trip cost on the y -axis against flow on the x -axis. The upward-sloping portion of the curve corresponds to congested travel; the downward-sloping portion corresponds to hypercongested travel. In the literature, this curve is referred to as the user cost curve or the supply curve of travel. The trip demand curve relates the (flow) demand for travel to trip price. Assume that no toll is applied, so that trip price equals user cost, and trip demand can be expressed as a function of user cost. Now draw in a linear trip demand curve that intersects the user cost curve three times, once on the upward-sloping portion of the user cost curve and twice on the backward-bending portion of the user cost curve. The first intersection point is a congested equilibrium, the latter two are hypercongested equilibria. Label the three equilibria e_1 , e_2 , and e_3 .

The issue that has been much debated concerns the stability of the latter two equilibria. Suppose, for the sake of argument, that an equilibrium is defined to be stable if, when an extra car is added to the entry traffic flow, the traffic flow eventually returns to that equilibrium's level. Even if the traffic inflow rate, apart from the added car, is held constant, solving for the transient dynamics of traffic flow is very difficult. But perhaps one should also take into account that the added car will affect traffic flow, hence user cost, and hence the traffic inflow rate in the future, which makes the analysis even more difficult. To circumvent this complexity, Else (1981) and Nash (1982), viewing equilibrium as the intersection of demand and supply curves, apply conventional economic adjustment dynamics without reference to the physics of traffic flow. Assuming that the addition of a car results in a disequilibrium increase in trip price and that disequilibrium adjustment occurs via price, Else argues that e_3 is locally stable. Assuming instead that the addition of a car results in a disequilibrium increase in flow and that disequilibrium adjustment occurs via quantity, Nash argues

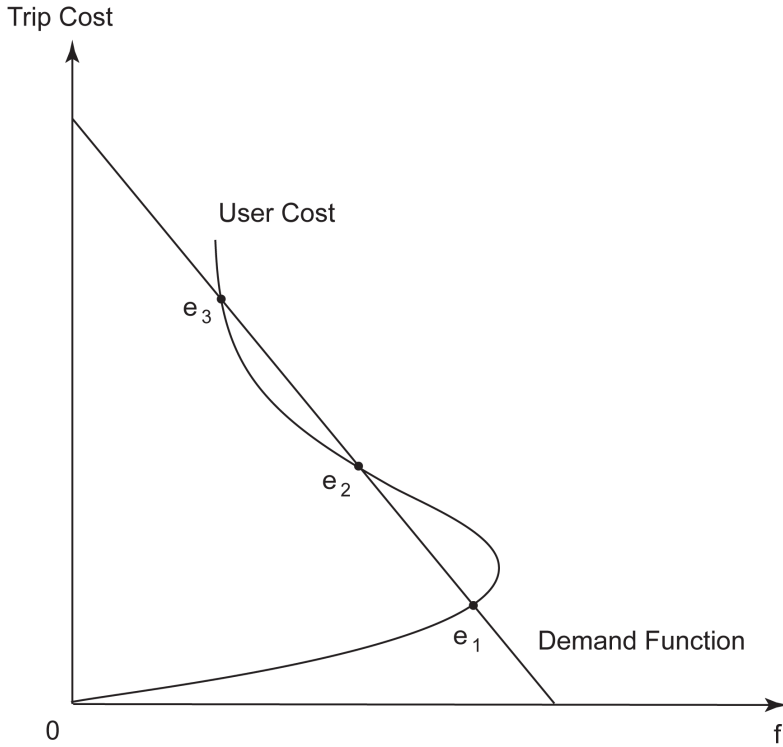


Figure 2: Stability of equilibria

that e_3 is locally unstable.³

There is now broad agreement that this stability issue cannot be resolved without dealing explicitly with the dynamics of traffic flow. Unfortunately, providing a complete solution even for traffic flow on a uniform point-input, point-output road with an exogenous inflow function is formidably difficult.⁴ The literature has responded in four qualitatively different ways to the intractability of obtaining complete solutions to this class of problems:

1. Derive qualitative solution properties, while fully respecting the physics of traffic

³Both papers consider only the case where the demand curve is flatter than the backward-bending portion of the user cost curve, and therefore do not investigate the properties of e_2 . Applying Else's analysis to e_2 would lead to the conclusion that it is unstable, applying Nash's that it is stable. Applying either analysis to e_1 would lead to the conclusion that it is stable.

⁴One inserts an equation relating velocity to density such as (i) into the equation of continuity (the continuous version of the conservation of mass), which yields a non-linear, first-order, partial differential equation. Applying the appropriate boundary conditions, one can in principle solve for density as a function of time and location along the road. Unfortunately, the partial differential equation does not have a closed-form solution for any sensible equation relating velocity and density.

flow.⁵ This approach is the ideal but is mathematically demanding.

2. Employ an assumption that simplifies the congestion technology, while continuing to treat location and time as continuous. One example is the “zero propagation” assumption that a car’s travel time on the road depends only on either the entry rate to the road at the time the car enters the road (Henderson, 1981) or the exit rate from the road at the time the car exits the road (Chu, 1995). Another example is the “infinite propagation” assumption that the speed of all cars on the road at a point in time depends on either the entry rate to the road or the exit rate from it (Agnew, 1977). Neither of these assumptions is consistent with classical flow theory. The question then arises as to whether the qualitative results of a model employing such assumptions are spurious.

3. Replace the partial differential equation with a discrete approximation – discretizing time and location – and then solve the resulting difference equation numerically. One such discrete approximation is Daganzo’s cell transmission model (Daganzo, 1992). Again, there is the concern that such approximations may give rise to spurious solution properties.

4. Adopt an even simpler traffic geometry in which the road system is isotropic, so that the entry and exit rates, as well as travel velocity, density, and flow, are the same everywhere on the network. This eliminates the spatial dimension of congestion so that the partial differential equation reduces to an ordinary differential equation. The second model of Small and Chu (2003) adopts this simplification, as do we in this paper. Unlike the previous two approaches, this approach does not entail any dubious approximation, but one may reasonably question the generality of results derived from models of an isotropic network.

Whatever approach is adopted, the issues arise as to the appropriate definition of equilibrium and the appropriate concept of stability to apply. This paper considers only steady-state equilibrium, in which the inflow rate and traffic flow remain constant

⁵Lindsey (1980) considered an infinite road of uniform width subject to classical flow congestion, with no cars entering or leaving the road. He proved that, if there is hypercongestion at no point along the road at some initial time, there will be hypercongestion at no point along the road in the future. Verhoef (1999) considered a finite road of uniform width with a single entry point and a single exit point. He argued (Prop. 2b) that if there is hypercongestion at no point along the road at some initial time, and if the constant inflow rate is less than capacity, there will be hypercongestion at no point along the road in the future. Verhoef (2001) developed the argument further using a simplified variant of car-following theory in which drivers control their speeds directly.

over time.⁶ The most familiar concept of stability is local stability. Start in a steady-state equilibrium. Perturb it (which implies an infinitesimal change) in some specified way. If the system always returns to that steady-state equilibrium, it is said to be locally stable with respect to that type of perturbation. In their textbook discussion of the stability of steady-state traffic equilibrium, Small and Verhoef employ a somewhat different concept of stability – dynamic stability. They define a stationary-state traffic equilibrium to be dynamically stable if it can arise as the end state following some transitional phase initiated by a change in the inflow rate.

The stage is now set to present Small and Verhoef’s textbook (2007, ps. 84-86) discussion of the stability of steady-state equilibria on a uniform road. They argue as follows:

“One difficulty with [the] conventional stability analysis [of steady-state traffic equilibria] is that the perturbations considered involve simultaneous change in the flow rates *into* and *along* the road, which is physically impossible. It therefore seems more appropriate to consider perturbations of the inflow rate, treating flow levels along the road as endogenous. Doing so introduces the concept of dynamic stability: can a given stationary state arise as the end state following some transitional phase initiated by a change in the inflow rate?”

Verhoef (2001) examines dynamic stability using the car-following model [...], allowing for vertical queuing before the entrance when inflows cannot be physically accommodated on the road. He finds that the entire

⁶Small and Chu (2003), Verhoef (2003), and Small and Verhoef (2006) consider two different types of equilibria. One type is the steady-state equilibrium, which was the focus of the earlier literature and will be the focus of this paper. Small and Verhoef reasonably question the relevance of these equilibria since practically traffic demand is not stationary over the day; rather, there are well-defined morning and evening rush hours. They examine an alternative concept of equilibrium that was originally proposed by Vickrey (1969) in the context of the bottleneck model (in which congestion takes the form of a vertical queue behind a bottleneck of fixed flow capacity). The idea is that over the rush hour individuals choose between traveling at an inconvenient time when traffic is less congested or at a convenient time when traffic is more congested. The full trip price then includes both inconvenience and travel time costs. With identical individuals, the equilibrium condition is that no individual can reduce her full trip price by altering her travel time. Verhoef (2003) refers to the corresponding equilibria as dynamic, while Small and Chu (2003) refers to it as an endogenous scheduling equilibrium. Employing this equilibrium concept, both papers argue that hypercongestion can occur in equilibrium over a portion of the rush hour, but not over the rush hour as a whole since equilibrium trip cost increases with the number of travelers.

hypercongested branch of the [user cost] curve is dynamically unstable. [...] [What arises instead is a dynamically stable steady state involving] a maximum flow on the road, a constant-length queue before its entrance, [...] and rates of queue-entries and queue-exits both equal to the capacity of the road. It does not involve hypercongestion on the road itself; rather, hypercongestion exists only within the entrance queue when it is modeled horizontally, as in Verhoef (2003). [...] Note that the flow rate and speed inside the queue are irrelevant to total trip time, making the economic properties of the model independent of the shape of the hypercongested portion of the speed-flow curve, even though traffic in the queue travels at a hypercongested speed.”

Essentially they argue that, when steady-state demand for the road is so high that its use cannot be rationed through congested travel, equilibrium exists, is unique and dynamically stable, and entails a steady-state queue at the entry point whose length adjusts to clear the market, with the road operating at capacity. In line with this argument, they replace the backward-bending portion of the user cost curve with a vertical segment at capacity flow.

In their papers, Small and Verhoef consider a variety of different models, including a couple in which the road is non-uniform. The properties of these models are broadly consistent with their textbook discussion. While Small and Verhoef do discuss contrary views and are not dogmatic, the *impression* left from their textbook discussion, complemented by their papers, is that, whatever the nature of the road system, the backward-bending portion of the steady-state user cost curve should always be replaced by a vertical segment at capacity flow, and that stable, steady-state hypercongested equilibria never exist. We suspect that their views are more nuanced than their textbook discussion suggests, but nonetheless that they would still argue that in normal circumstances their proposed resolution of the debate applies.

In this paper, we shall undertake an exhaustive stability analysis of a simple model in which stable, steady-state hypercongested equilibria do exist (by exhaustive stability analysis, we mean one that provides a complete characterization of the model’s transient dynamics with steady-state demand from all initial conditions). The model is of an isotropic road network. Its essential feature is that cruising for parking occurs

in equilibrium, with cars cruising for parking behaving like a (random access) queue that interferes with traffic flow. While a particular model does not prove that Small and Verhoef’s resolution is less general than they appear to believe, in the concluding discussion we argue that the reduction in throughput caused by cruising for parking is representative of a wide range of phenomena in heavily-congested traffic systems and of congestible systems in general.

3 Model Description

The model is aimed at describing downtown traffic and its interaction with on-street parking. A detailed description of a slightly different version of the model, which focuses only on the steady states under saturated parking conditions, can be found in Arnott and Inci (2006). This paper treats the nonstationary dynamics of the model with a special focus on stability and allows for transitions between saturated and unsaturated parking conditions.

The downtown area has an isotropic (spatially homogeneous) network of streets. For concreteness, one can imagine a Manhattan network of one-way streets. We assume that all travel is by car and that there are only on-street parking spaces.⁷ Each driver enters the downtown area, drives to his destination, parks there immediately if a vacant parking space is available and otherwise circles the block until a parking space becomes available, visits his destination for an exogenous length of time, and then exits the downtown area. Drivers differ in driving distance and visit length. Driving distance is Poisson distributed in the population with mean m , and visit length is Poisson distributed with mean l .

Downtown parking spaces are continuously provided over the space. Drivers are risk neutral expected utility maximizers. There may be three kinds of cars on streets: *cars in transit*, *cars cruising for parking*, and *cars parked*. Apart from the architecture of the street (*i.e.*; a curbside allocated to parking and the width of the streets), the travel speed depends on the density of these two types of cars on the streets. We assume that cars cruising for parking slow down traffic more than cars in transit and

⁷Arnott (2006) focuses on off-street parking in a downtown area and Arnott and Rowse (forthcoming) extends the current model to allow for both on- and off-street parking.

thus contribute more to congestion.

Let T be the pool of cars in transit per unit area, C be the pool of cars cruising for parking per unit area, and P be the pool of on-street parking spaces per unit area (which is held constant throughout the paper). The traffic technology is defined by an in-transit travel time function $t(T, C, P)$ where t is per unit distance.⁸ Let P_{\max} be the maximum possible number of on-street parking spaces per unit area. We assume that the technology satisfies $t_T > 0$, $t_C > 0$, $t_P > 0$, $t(0, 0, P) > 0$, $\lim_{P \rightarrow P_{\max}} t(T, C, P) = \infty$, and t is convex in T , C , and P .

Denoting the rate of entry into the network per unit area-time by D and the exit rate from the pool of cars in transit by E , we can write the rate of change in the pool of cars in transit as follows:

$$\dot{T}(u) = D(u) - E(u) \quad , \quad (1)$$

where u is the time. This trivially describes the evolution of the pool of cars in-transit at every instant. Describing the evolution of downtown parking is less trivial. If the amount of curbside parking constrains the flow of cars the traffic system can accommodate, there are two régimes of downtown parking and the system may switch from one to the other.

In the first régime, the downtown parking is *saturated*, meaning that a vacant parking space is immediately taken by a car cruising for parking. In this régime, all parking spaces are filled at any given time but the pool of cars cruising for parking evolves over time. Therefore, when parking is saturated, the rate of change in the pool of cars cruising for parking is simply the difference between the entry rate into the pool of cars cruising for parking and the exit rate from it, or simply

$$\dot{C}(u) = E(u) - Z(u) \quad , \quad (2)$$

where E now denotes the entry rate into the pool of cars cruising for parking from the pool of cars in transit, which equals the exit rate from the in-transit pool, and Z the exit rate from the pool of cars cruising for parking. In this régime, the pool of occupied parking spaces, S , remains fixed (so that $\dot{S} = 0$), but the pool of cars

⁸Note that we assume that P enters the in-transit travel time function even when parking is unsaturated. The rationale is that even one car parked curbside on a city block precludes the use of that lane for traffic flow over the entire block.

cruising for parking evolves.

In the second régime, parking is *unsaturated*, meaning that there are empty parking spaces so that cars in transit can find a parking space upon arrival at their destinations.⁹ In this régime, the stock of cars cruising for parking is zero (so that trivially \dot{C} is zero too) but the pool of occupied parking spaces evolves. The evolution of S is given by

$$\dot{S}(u) = E(u) - X(u) \quad , \quad (3)$$

where E is now the entry rate into the pool of occupied parking spaces from the pool of cars in transit, which equals the exit rate from the in-transit pool, and X the exit rate from the pool of occupied parking spaces.

To be able to write the equations of motion more specifically, we shall now turn to characterizing D , E , Z and X in detail. The demand function, D , is a function of the perceived mean full trip price, F ,

$$D = D(F); \quad D(0) = \infty, \quad D(\infty) = 0, \quad D' < 0 \quad . \quad (4)$$

The mean full trip price at time u depends on the in-transit travel time cost, the (mean) cruising-for-parking time cost, and the cost of on-street parking, at this point in time. Denoting the value of time with ρ and the on-street parking fee with λ , the mean full trip price can be written as follows¹⁰:

$$F = \rho(mt(T(u), C(u), P) + \frac{C(u)l}{P}) + \lambda l \quad . \quad (5)$$

The exit rate from the in-transit pool equals the stock of cars in the in-transit pool multiplied by the probability that a car will exit the in-transit pool per unit time¹¹:

$$E(u) = \frac{T(u)}{mt(T(u), C(u), P)} \quad . \quad (6)$$

⁹Arnott and Rowse (1999) provides a more sophisticated treatment of unsaturated parking in which cruising for parking occurs. In contrast to the model of this paper, their city is located on an annulus. On the basis of the parking occupancy rate, a driver decides how far from his destination to start cruising for parking, takes the first available vacant space, and walks from there to his destination. Adapting this more sophisticated treatment of unsaturated parking here should not substantially alter our results.

¹⁰One could define the full price of a trip to include the time cost of a visit, as is done in Arnott and Inci (2006). λ would then be defined as the time and money cost of a visit per unit time.

¹¹Appendix A.2 derives this equilibrium condition.

Due to the assumption that visit durations are generated by a Poisson process, the probability that an occupied parking space is vacated per unit time is $1/l$. Thus, when parking is saturated, the exit rate from the cruising-for-parking pool equals that probability multiplied by the number of parking spaces, P :

$$Z(u) = \frac{P}{l} \quad . \quad (7)$$

When parking is unsaturated, the exit rate from the pool of occupied parking spaces is defined similarly. X is the probability that a particular parking space is vacated, $1/l$, times the number of occupied parking spaces at that particular time, $S(u)$:

$$X(u) = \frac{S(u)}{l} \quad . \quad (8)$$

After substituting out the variables E , Z , and X , downtown traffic is characterized by the following autonomous differential equation system with two régimes.

$$\text{Régime 1} : \begin{cases} \dot{T}(u) = D(\rho(mt(T(u), C(u), P) + \frac{C(u)l}{P}) + \lambda l) - \frac{T(u)}{mt(T(u), C(u), P)} \\ \dot{C}(u) = \frac{T(u)}{mt(T(u), C(u), P)} - \frac{P}{l} \\ \dot{S}(u) = 0 \end{cases} \quad (9)$$

$$\text{Régime 2} : \begin{cases} \dot{T}(u) = D(\rho(mt(T(u), 0, P) + \lambda l) - \frac{T(u)}{mt(T(u), 0, P)} \\ \dot{C}(u) = 0 \\ \dot{S}(u) = \frac{T(u)}{mt(T(u), 0, P)} - \frac{S(u)}{l} \end{cases} \quad . \quad (10)$$

In Section 4, we shall focus on these two régimes in turn. That the two differential equation systems are autonomous allows us to employ a phase plane analysis to investigate the stability of the traffic system, converting what would otherwise be an essentially intractable problem into one that is straightforward to analyze. To achieve “autonomy”, we made three essential simplifying assumptions: *i*) trip length is Poisson distributed; *ii*) visit duration is Poisson distributed; and *iii*) travel demand at time u is a function only of the state variables, C and T , at time u . The former Poisson assumption makes the exit rate from the pool of cars at time u dependent on only the stock of cars in transit and cruising for parking at time u . The latter Poisson assumption makes the exit rate from the pool of parked cars at time u dependent only

on the stock of cars parked at time u . The three assumptions together imply that the dynamics of the traffic system depend only on the system's state variables, T , C , and S , and not separately on time. Put alternatively, the history of the traffic system is fully captured by the values of the state variables.

None of these assumptions is realistic. The assumption on demand is particularly objectionable because it is hard to justify on the basis of microfoundations. Our justification for making these assumptions is that together they permit a complete stability analysis that fully respects the physics of traffic flow, which no previous papers have been able to perform. Thus, our paper provides a rigorous stability analysis of a highly particular model. While we cannot claim that the results of our analysis are general, we believe that the underlying reasons that the stability properties of the model differ from those of Small and Verhoef are general. In particular, we believe that our model is representative of traffic systems in which the dissipative activity required to clear the travel market when demand is high undermines system performance.

4 Analysis of Steady-state Equilibrium

In this section, we characterize the steady-state equilibria of the model and display them graphically. In any steady-state equilibrium, the entry rate into each pool equals the corresponding exit rate from it, so that the size of each pool is time invariant. We have the following definitions:

Definition 1 (Saturated equilibrium) *A saturated steady-state equilibrium is a triple $\{T, C, S\}$ such that $\dot{T}(u) = 0$, $\dot{C}(u) = 0$, $\dot{S}(u) = 0$, and $S = P$.*

Definition 2 (Unsaturated equilibrium) *An unsaturated steady-state equilibrium is a triple $\{T, C, S\}$ such that $\dot{T}(u) = 0$, $\dot{C}(u) = 0$, $\dot{S}(u) = 0$, and $C = 0$.*

4.1 Régime 1: Saturated steady-state equilibria

We shall start with the saturated steady-state equilibrium associated with régime 1 shown in (9). There are cars cruising for parking in any traffic equilibrium in which

parking is saturated. The parking spots are completely full at any given time and once a spot is vacated it is immediately filled by a car that is currently cruising for parking. We make two additional assumptions regarding the traffic technology and the street architecture. First, we assume that cars cruising for parking contribute to congestion more than cars in transit.

Assumption 1 $t_C > t_T$.

We distinguish between flow and throughput. We define flow to be the number of cars passing a point on a street per unit time. Flow therefore includes both cars in transit and cars cruising for parking. By multiplying the flow, so defined, by the number of streets per unit area, we could define flow per unit area. We define throughput to be the entry rate of cars into a unit area, which in steady state equals the exit rate of cars per unit area. Since cars cruising for parking circle around the block rather than enter or exit the downtown area, throughput includes only cars in transit. Thus cars cruising for parking do not contribute to throughput but only to the traffic flow. Throughput capacity, the maximum throughput consistent with the congestion technology, which occurs when there are no cars cruising for parking, equals $\max_T \{T \div (mt(T, 0, P))\}$. The second assumption is that this exceeds the exit rate from saturated parking, P/l , since otherwise parking would never be saturated in a steady-state equilibrium.

Assumption 2 $\max_T \frac{T}{mt(T,0,P)} > \frac{P}{l}$.

These two assumptions along with the assumption of convexity of t in T and C imply that $T/(mt(T, 0, P)) = P/l$ has two roots. For the existence of a saturated traffic equilibrium, the entry rate in the absence of cruising for parking must lie between these two roots. Arnott and Inci (2006) proved that there is a unique saturated steady-state equilibrium when that holds. Apart from the fact that the pool of occupied parking spaces is time invariant with $\dot{S}(u) = 0$, the unique equilibrium is characterized by two equations. First, we know that, in any saturated steady-state equilibrium, the entry rate into the in-transit pool equals the exit rate from that pool,

$$D = \frac{T}{mt} \quad . \quad (11)$$

Second, with saturated parking, the entry rate into the cruising-for-parking pool equals the exit rate from it,

$$\frac{T}{mt} = \frac{P}{l} . \quad (12)$$

Eqs. (11) and (12) represent $\dot{T}(u) = 0$ and $\dot{C}(u) = 0$, respectively.¹² Figure 3 draws these equations in the $T - C$ space with plausible functional specifications taken from Arnott and Inci (2006) that we specify below.¹³

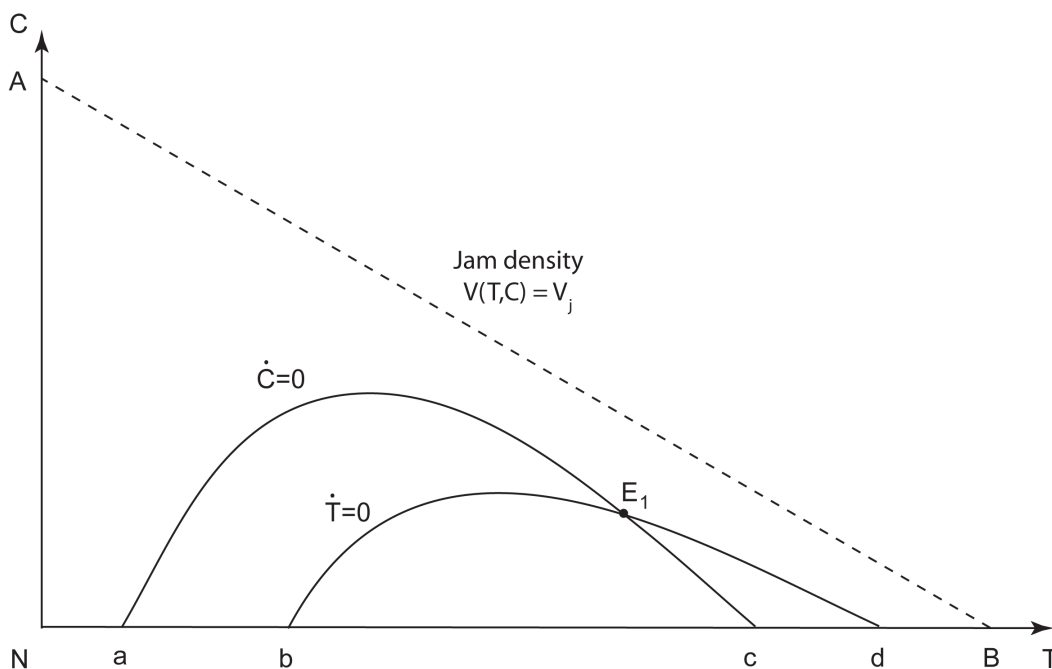


Figure 3: Saturated, steady-state equilibrium in $T - C$ space

Suppose that travel time t is weakly separable between (T, C) and P ; refer to the sub-function $V(T, C)$ as the *effective density function*, and $\omega(P)$ as the *effective capacity function*. As usual, suppose also that t depends on the ratio of effective density and capacity, so that $t = t(V(T, C)/\omega(P))$. We measure effective density in terms of

¹²A quick way to see the uniqueness of the equilibrium in $T - C$ space is to make use of the steady-state relationship, $mt = Tl/P$, which yields $D = P/l$. Then the two equations governing the stationary equilibrium are $D = P/l$ and $T/(mt) = P/l$, which intersect once, if they intersect.

¹³Figure 3 draws bE_1d as the $\dot{T} = 0$ locus. Technically, there is another portion of the $\dot{T} = 0$ locus, the jam density line. Since traffic is jammed, the trip price is infinite so that the demand inflow is zero, and the trip time is infinite so that T/mt is also zero. Since this portion of the locus is irrelevant in the stability analysis, we shall refer to bE_1d as the $\dot{T} = 0$ locus in this space.

in-transit car equivalents, and assume it to take the following form:

$$V(T, C) = T + \theta C; \quad \theta > 1 \quad , \quad (13)$$

so that a car cruising for parking contributes θ times as much to congestion as a car in transit. Finally, we assume that Greenshield's Relation (1935) holds, so that the speed of cars is a decreasing linear function of effective density. We therefore have

$$t = \frac{t_0}{1 - \frac{V(T, C)}{V_j}} \quad , \quad (14)$$

where t_0 is free-flow travel time and V_j is jam density. We shall also assume that demand is iso-elastic so that

$$D(F) = D_0 F^a \quad , \quad (15)$$

where $D_0 > 0$ is a measure of the market size and $a < 0$ the constant elasticity of demand. Given these assumptions, as shown in Figure 3, the implicit function $C(T)$ defined by $\dot{C}(u) = 0$ is a concave function having two roots at $C = 0$, both of which are greater than zero and less than V_j . The $\dot{T} = 0$ locus has two parts. The first intersects $C = 0$ potentially multiple times between zero and less than V_j , The second, which is not immediately obvious, is the jam density line, since at jam density, the LHS of (11) is zero since the trip price is infinite and the RHS is zero since trip time is infinite. As mentioned above, if the $\dot{C}(u) = 0$ and $\dot{T}(u) = 0$ loci intersect they do so once, establishing the unique saturated equilibrium, E_1 , shown in Figure 3. In section 4.3, we shall define congestion and hypercongestion. According to the definitions there, whether E_1 is congested or hypercongested depends on where the $\dot{C} = 0$ and $\dot{T} = 0$ loci intersect. The "qualitative" curvature of the figures in this paper can be obtained with the following parametric specifications: $m = 2$ miles, $l = 2$ hours, $\rho = \$20$ per hour, $t_0 = 0.05$ hours per mile, $P = 3712$ parking spaces per square mile, $V_j = 1778.17$ per square mile, $\lambda = \$1$ per hour, $D_0 = 3190.04$, and $a = -0.2$.

For future reference, note that the $\dot{C} = 0$ locus cuts the T -axis at points a and c , and that with the assumed functional forms, the $\dot{T}(u) = 0$ locus cuts the T -axis twice, at points b and d . One other important thing to mention is that there can be

no equilibrium above the jam density line AB ($T + \theta C = V_j$).¹⁴ Hence, the relevant subspace for the analysis of saturated equilibria is inside the triangle ANB (where N is the point where $C = 0$, $T = 0$)¹⁵. We shall see that under Assumption 2, which states that the maximum throughput is limited by the parking constraint, there are two unsaturated equilibrium, in addition to the saturated equilibrium, one of which corresponds to gridlock. If Assumption 2 did not hold, the parking constraint would never bind, and there would be three unsaturated equilibria.

4.2 Régime 2: Unsaturated steady-state equilibria

Unsaturated equilibria correspond to régime 2 whose equation system is given in (10). The stock of cars cruising for parking is zero so that a driver finds a parking space immediately upon reaching his destination. The stock of occupied parking space adjusts until the system reaches a steady state. Apart from $C(u) = 0$ (and $\dot{C}(u) = 0$), two equations characterize an unsaturated steady-state equilibrium. The first is again that the entry rate into the in-transit pool equal the exit rate from it:

$$D = \frac{T}{mt} \quad . \quad (16)$$

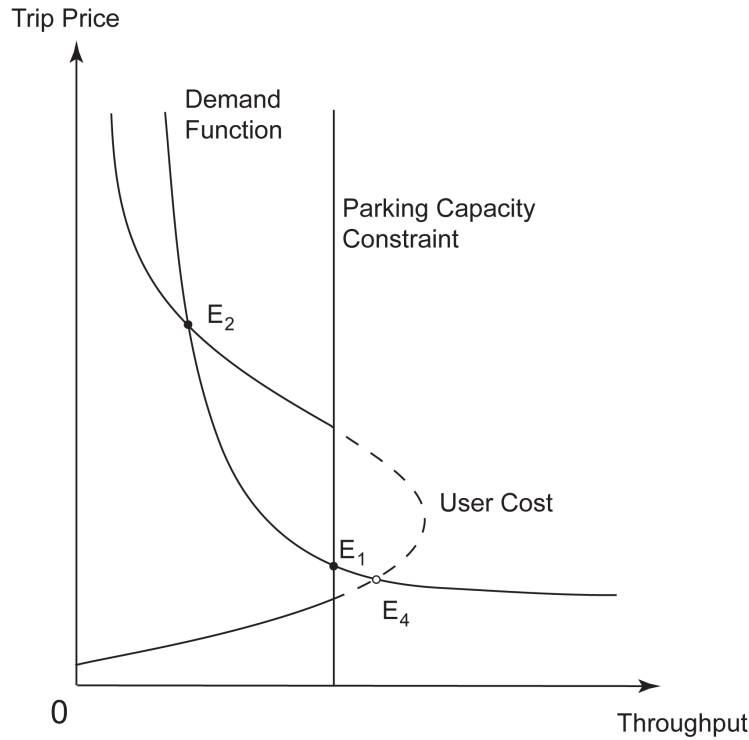
The second is that the entry rate into the pool of occupied parking spaces equal the exit rate from it.

$$\frac{T}{mt} = \frac{S}{l} \quad . \quad (17)$$

Figure 4 draws these equations in $T - S$ space with the functional specifications indicated above. Eq. (16) is the same as (11) with $C = 0$. Thus, in $T - S$ space one part of the $\dot{T} = 0$ locus is vertical at the T coordinates corresponding to the points b and d , the other part is vertical at jam density. Eq. (17) has an inverted U-shape, passes through the origin and $(V_j, 0)$, and intersects $S = P$ at the points a and c , where the $\dot{C} = 0$ locus intersects $C = 0$. Thus, each of the vertical lines associated with $\dot{T} = 0$ intersects $\dot{S} = 0$ exactly once, leading to three potential unsaturated equilibria, E_2 , E_3 , and E_4 .

¹⁴Given (13), the jam density curve is linear as shown in Figure 3, but not generally otherwise.

¹⁵Later we work in (T, C, S) space, for which the origin is $(0, 0, 0)$. We do not refer to the point N as the origin since its coordinates in this space are $(0, 0, P)$.



Note: There is another equilibrium, E_3 , in which flow is zero and trip price is infinite.

Figure 5: Unsaturated, steady-state equilibrium in flow-trip price space

properties of the three equilibria.

4.3 Identifying hypercongestion

Recall that we have made a distinction between the (physical) density of traffic measured in cars per unit area, and the effective density of traffic measured in in-transit car-equivalents per unit area, which takes into account that a car cruising for parking generates more congestion than a car in transit. The fundamental identity of traffic flow holds if flow and density are both defined in terms of physical cars. It also holds if flow and density are defined in terms of car equivalents. We have however chosen to work with effective density, since that is what traffic congestion is a function of, but to use the term flow to refer to the physical flow of cars, since that is what a bystander would observe. Thus, we must proceed with care.

We have assumed that traffic congestion is described by Greenshield's Relation,

adapted to take into account cars cruising for parking. In particular, we have assumed that

$$v = v_f \left(1 - \frac{V}{V_j}\right) \quad \text{or} \quad V = V_j \left(1 - \frac{v}{v_f}\right) \quad , \quad (18)$$

where v is velocity, v_f free flow velocity, V is effective density, and V_j is effective jam density. We define congestion and hypercongestion in the following way:

Definition 3 (Congestion, hypercongestion) *Congestion occurs when traffic velocity is greater than that associated with capacity throughput, hypercongestion when traffic velocity is less than that associated with capacity throughput.*

Since capacity throughput is the maximum possible throughput, which occurs when there are no cars cruising for parking, its calculation does not require distinguishing between effective and physical cars. Capacity throughput equals capacity flow:

$$f^c = \max_v vV(v) = \max_v v \left(V_j \left(1 - \frac{v}{v_f}\right) \right) = \frac{V_j v_f}{4} \quad . \quad (19)$$

The velocity associated with capacity throughput is $v_f/2$. Since there is a one-to-one correspondence between velocity and effective density, we may equivalently define traffic to be hypercongested if effective density is greater than that associated without capacity throughput. Thus, we have that travel is hypercongested if

$$V(T, C) > \frac{V_j}{2} \quad , \quad (20)$$

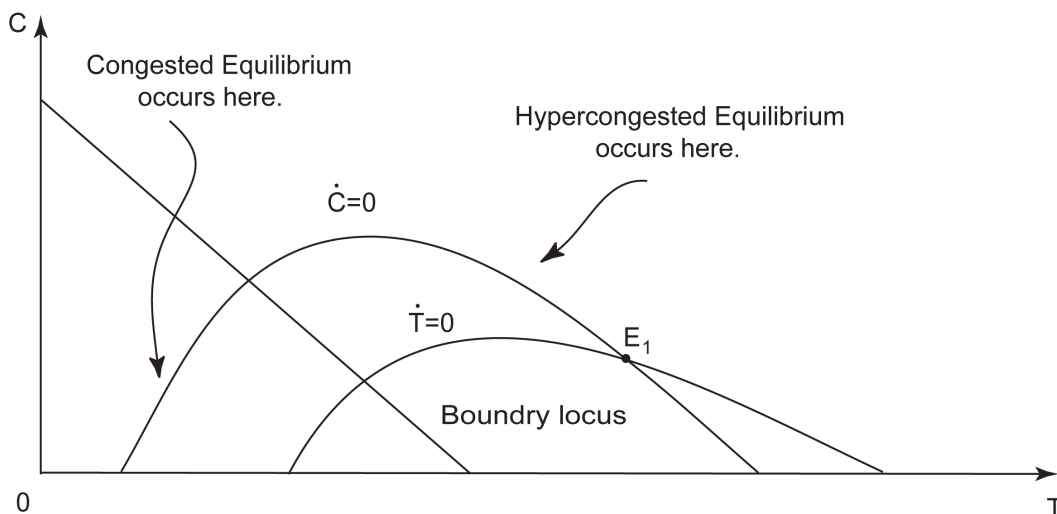
and congested when the inequality is reversed. For the particular effective density function we have assumed in (13), we obtain that travel is hypercongested if

$$T + \theta C > \frac{V_j}{2} \quad , \quad (21)$$

and is congested otherwise. We refer to the equation $T + \theta C = V_j/2$ as the *boundary locus*, since it separates the region of congested travel from the region of hypercongested travel. Figure 6 plots the boundary locus, as well as the $\dot{T} = 0$ and $\dot{C} = 0$ loci in $T - C$ space. The boundary locus has a slope of $-1/\theta$. Travel below the locus is congested, and above the locus is hypercongested. We define equilibria to be congested or hypercongestion accordingly. In particular:

Definition 4 (Congested equilibrium, hypercongested equilibrium) *An equilibrium is congested when congestion according to Definition 3 occurs, and hypercongested otherwise.*

As drawn, the equilibrium E_1 is hypercongested.



Note: The figure is drawn choosing parameters such that the saturated equilibrium is hypercongested. With a different choice of parameters the saturated equilibrium can instead be congested.

Figure 6: Identifying hypercongested travel in $T-C$ space

5 Stability Analysis

This section carries out the stability analysis by combining the two régimes. We start our analysis by stating our notions of hypercongestion and (dynamic) stability.

Definition 5 (Stability) (i) A steady-state equilibrium is said to be **locally stable**¹⁷ if it can be reached from all initial traffic conditions in its neighborhood; (ii) a steady-state equilibrium is said to be **saddle-path stable** if it can be reached only from initial traffic conditions on one of its arms; (iii) A steady-state equilibrium is said to

¹⁷This is sometimes called *asymptotically stable*.

be **dynamically stable** if it can be reached from at least one initial traffic condition other than itself.¹⁸

For a complete stability analysis, we need to take into account not only transition between the two régimes but also the possibility that traffic might get stuck at jam density. Régime 1 (the saturated régime) is shown in Figure 3 in $T - C$ space, and régime 2 (the unsaturated régime) in Figure 4 in $T - S$ space. The two régimes may be analyzed simultaneously in the three-dimensional figure in $T - C - S$ space displayed in Figure 7. We shall explain how to read this figure before analyzing the stability of the equilibria.

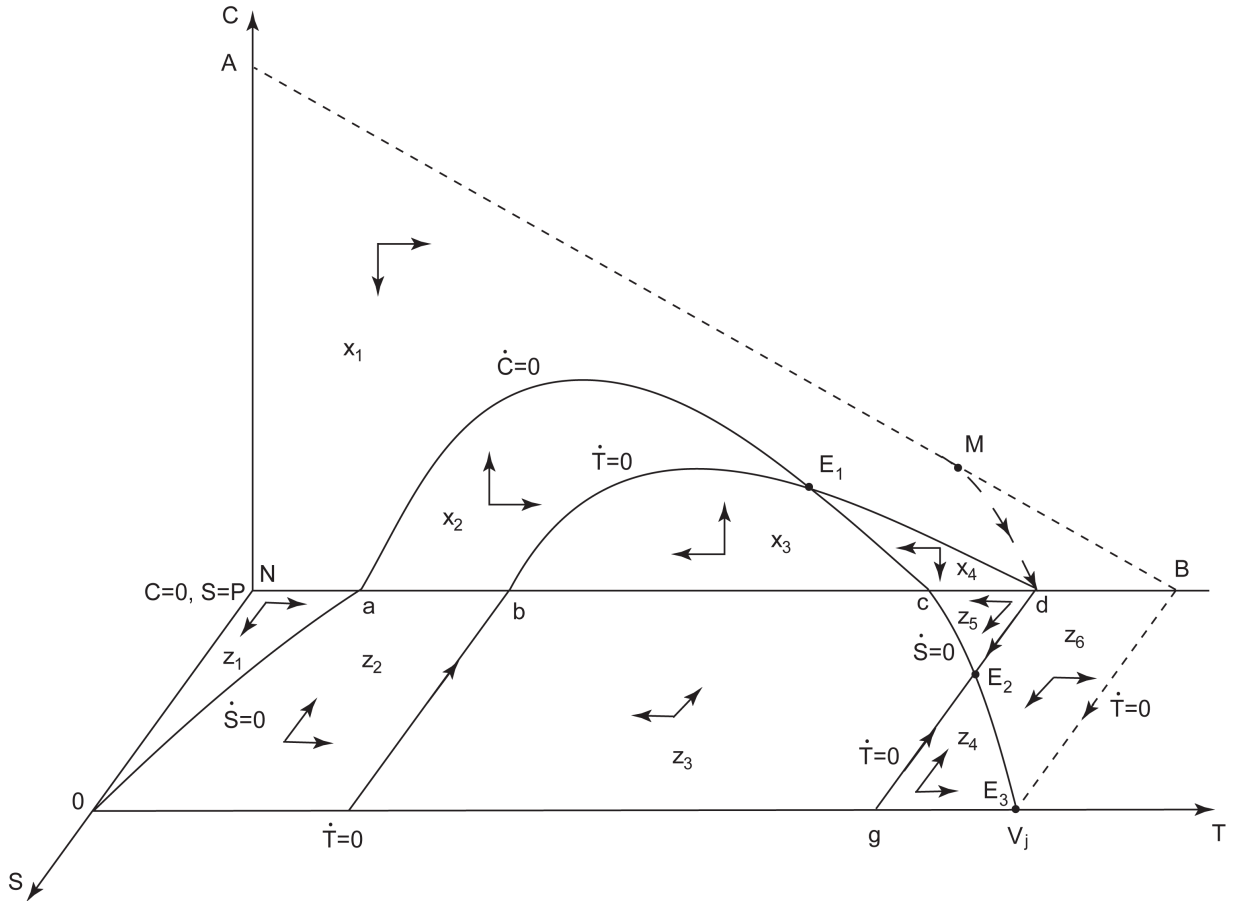


Figure 7: Saturated and unsaturated steady-state equilibrium in $T-C-S$ space

The vertical $T - C$ plane reproduces Figure 3 with some added detail. The horizontal $T - S$ plane reproduces Figure 4 with some added detail. The fold where the two planes

¹⁸Dynamic stability is therefore weaker than the other two.

join is along $C = 0$ and $S = P$, with N representing the point $(T, C, S) = (0, 0, P)$ and B the point $(V_j, 0, P)$.

Consider the $T - C$ plane. The line AB corresponds to jam density. Since densities above jam density are infeasible, the feasible region of the plane is the triangle NAB . The $\dot{T} = 0$ and the $\dot{C} = 0$ loci divide the plane into four areas, labeled x_1, x_2, x_3 , and x_4 . In each of these regions, the direction of motion of C and T – shown by the arrows – is the same; for example, in region x_1 , C is decreasing and T is increasing. The point M is the point on the jam density locus whose trajectory leads to the point d .

Now consider the $T - S$ plane. The line BV_j corresponds to jam density. Since densities above jam density are infeasible, the feasible region of the plane is the rectangle $ONBV_j$. The $\dot{S} = 0$ locus and the three parts of the $\dot{T} = 0$ locus divide the plane into six areas, z_1, z_2, z_3, z_4, z_5 , and z_6 . In each of these regions, the direction of motion is the same; for example in region z_1 , T is increasing and S is decreasing.

In summary, x_i and z_j (where $i \in \{1, \dots, 4\}$ and $j \in \{1, \dots, 6\}$) denote areas, and the corresponding vector fields for each area are shown in each of them; the point 0 is the origin of the 3D figure; the points a, b, c , and d are as defined before; the dotted lines indicate jam density situations; E_1 is the saturated equilibrium; E_2 is an unsaturated equilibrium; E_3 is another unsaturated equilibrium in which there is a gridlock; as drawn, all of these equilibria are hypercongested.¹⁹

We shall now state the only proposition of the paper, from which we deduce our main findings.

Proposition 1 *Any starting point on the locus MdE_2g moves to E_2 . Any starting point to the left of the locus moves to E_1 , and any starting point to the right of the locus moves to E_3 .*

Proof. We shall prove this proposition in three steps.

1. Areas in the triangle ANB :

¹⁹One might argue the possibility of a limit cycle. However, it is ruled out by Bendixson's Nonexistence Criterion for the equations of motion of régime 1. Vector fields show that it cannot happen for the equations of motion of régime 2, either. We conjecture that there cannot be a limit cycle circling between the régimes. Figure 9 in Appendix A.1 shows couple of trajectories.

- **Area x_1 (excluding the adjustment path Md)** : The vector fields in this area point south-east. Any initial condition in x_1 will either hit E_1 or E_3 .
 - For sufficiently high values of C , the trajectories will reach the equilibrium E_1 . They will approach the $\dot{C} = 0$ locus before reaching E_1 since the vector fields right below the locus (in area x_2) point north-east.
 - The trajectories for low values of T and C will hit the line segment Na . Once they hit Na , C cannot further decrease since it cannot go below zero. Thus, parking becomes unsaturated and the vector fields in area z_1 will apply. The trajectory will pass through $a0$ and enter the area z_2 . The vector fields in this area will then carry the trajectories toward the line segment ab either via area z_2 or via the $\dot{T} = 0$ locus in the $T - S$ plane. On the line segment ab , S cannot further decrease since it has to be nonnegative. Thus, parking becomes saturated again. Then, the vector fields shown in x_2 will apply and therefore the trajectory will once again hit E_1 .
 - For some intermediate values of T , the trajectories may hit the curve E_1d and pass through the area x_4 . At this time, the trajectory may either hit E_1c and move into x_3 (and maybe x_2 after that) before reaching E_1 or it may hit the line segment cd . If it hits cd , parking will have to become unsaturated. Then, the trajectory moves into the area z_5 followed by area z_3 . Once it is in z_3 , the trajectory will move toward the line segment bc and then parking becomes saturated again before reaching E_1 from the area x_3 .
 - Yet another possibility occurs for sufficiently large values of T and sufficiently small values of C . There has to be an initial condition M such that the trajectory initiated from M passes through the point where the $\dot{T} = 0$ locus cuts the NB line, namely point d . Given that trajectories in this differential equation system cannot intersect unless one is on the same trajectory as the other, for any initial point on the right hand side of the path Md , the trajectory will hit the line segment dB . Once it hits there, C cannot further decrease, parking becomes unsaturated, and the trajectory will move into the area z_6 . Given the vector fields there, it is then obvious that the trajectory

will move towards E_3 to establish a gridlock of cars on the network of streets. The vector fields in z_6 cannot carry a trajectory towards E_2 .

- **Area x_2** : The vector fields in this area point north-east. Any trajectory from any initial condition in this area will trivially reach E_1 . Parking never becomes unsaturated along the adjustment path as C will increase at all times.
- **Area x_3** : The vector fields in this area point north-west. There are two possibilities in this area. For lower values of T , the trajectories will enter x_2 (or move along the border of x_2 and x_3) before reaching E_1 . For higher values of T , they hit the equilibrium E_1 from the area x_3 . Parking never becomes unsaturated along the adjustment path as C will continuously increase until it reaches a steady-state equilibrium according to the vector fields.
- **Area x_4** : The vector fields in this area point south-west. There are two possibilities in this region. First, the trajectory may hit E_1c and enter the area x_3 before hitting E_1 . The other possibility is that the trajectory may hit the line segment cd . Once it hits cd , parking becomes unsaturated. The trajectory then moves into the area z_5 followed by area z_3 before reaching E_1 , as previously explained.

2. Areas in the rectangle $ONBV_j$:

- **Area z_1** : The vector fields in this area point down-east. Therefore, any trajectory in this area hits the curve $a0$ and passes into the area z_2 . The vector fields in this area will then carry the trajectories toward the line segment ab either via area z_2 or via the $\dot{T} = 0$ locus. On the line segment ab , S cannot increase further since it cannot exceed P . Thus, parking becomes saturated. Then, the vector fields shown in x_2 will apply and therefore the trajectory will hit E_1 , as previously explained.
- **Area z_2** : The vector fields in this area point up-east. They will carry the trajectories toward the line segment ab either via area z_2 or via the $\dot{T} = 0$ locus. On the line segment ab , S cannot further decrease since it has to be nonnegative. Thus, parking becomes saturated and the vector fields of the area x_2 will apply. Consequently, the trajectory will reach E_1 .

- **Area z_3** : The vector fields in this area point up-west. Any trajectory in this area will hit bc and reach E_1 , as previously explained.
- **Area z_4** : The vector fields in this area point up-east. Any trajectory in this area will first hit E_2V_j and then follow this curve until it reaches the gridlock equilibrium E_3 .
- **Area z_5** : The vector fields in this area point down-west. As previously explained, any trajectory here will first hit cE_2 and then enter into z_3 before reaching E_1 , as previously explained.
- **Area z_6** : The vector fields in this area point down-east. Any trajectory here will either directly hit E_3 or follow BV_j before doing so.

3. Points on the locus MdE_2g :

- **Points on the line segment E_2g** : Since $\dot{T} = 0$ but $\dot{S} > 0$, any trajectory initiated on this line segment will follow the $\dot{T} = 0$ locus until it reaches E_2 .
- **Points on the line segment E_2d** : Since $\dot{T} = 0$ but $\dot{S} < 0$, any trajectory initiated on this line segment will follow the $\dot{T} = 0$ locus until it reaches E_2 . This, along with other parts of the proof, implies that E_2 is a saddle point.
- **Points on the adjustment path Md** : Any trajectory initiated from this curve will first hit d . However, since C cannot be negative, parking will become unsaturated. Thereafter, the trajectory will follow the $\dot{T} = 0$ locus until it reaches E_2 , as previously explained.

■

There are two important corollaries to Proposition 1.

Corollary 1 *The hypercongested saturated equilibrium E_1 , the hypercongested unsaturated equilibrium E_2 , and the hypercongested gridlock equilibrium E_3 are all dynamically stable. E_1 and E_3 are both locally stable equilibria. E_2 is not locally stable but is saddle-path stable.*

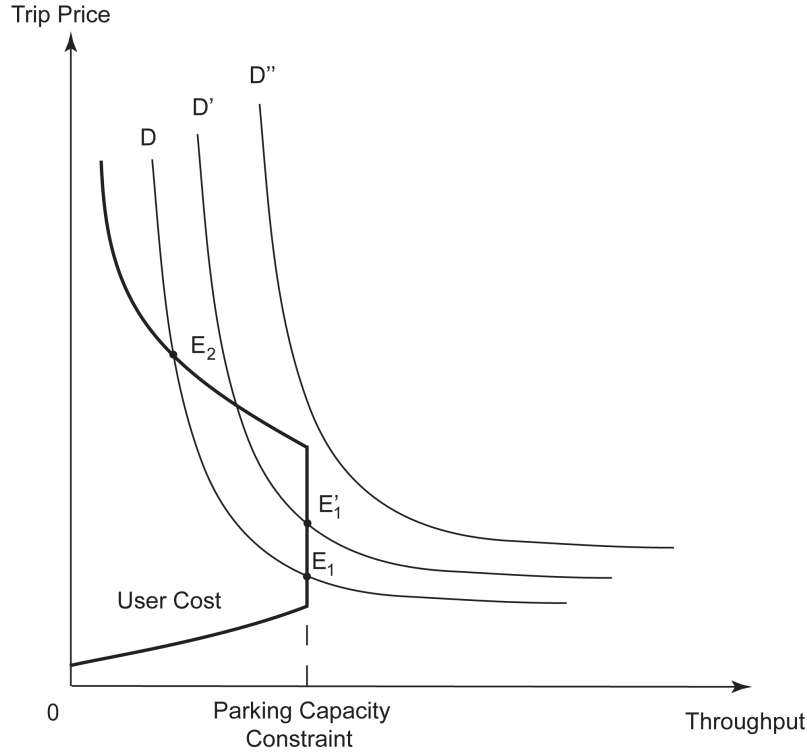
This corollary follows directly from Proposition 1 and the definition of dynamic stability in Definition 5.

Corollary 2 *There is no gridlock equilibrium with cruising for parking.*

The intuition for this result is straightforward. Start with a situation with traffic gridlock and cruising for parking. Since traffic is gridlocked, the exit rate from the in-transit pool and hence the entry rate into the cruising-for-parking pool is zero. Since parking is saturated, the exit rate from the cruising-for-parking pool is P/l . The cruising-for-parking pool therefore shrinks.

That completes our formal stability analysis. To provide some intuition for the traffic system's dynamics, we consider starting in area z_4 and investigate how traffic and parking adjust along the path to the jam density equilibrium E_3 . In area z_4 , travel is so slow that the exit rate from the in-transit pool is lower than the inflow, so that the size of the in-transit pool increases. Since S is low, the exit rate from the in-transit pool is still larger than the rate at which parking is vacated, so that S increases. Eventually, however, as travel gets slower and slower and the exit rate from the in-transit pool decreases, a point is reached where the exit rate from the parking pool equals the exit rate from the in-transit pool. As time proceeds, travel becomes even slower, the exit rate from the in-transit pool declines and falls short of the exit rate from the parking pool. The density of cars in transit continues to increase and the stock of parked cars decreases asymptotically towards the equilibrium E_3 . Even though E_3 cannot be reached from the origin with a time-invariant demand function, a demand pulse may push traffic into the regions z_4 or z_6 , or to the right of Md in the saturated régime, and once in those regions, with a time-invariant demand function, there is no way of escaping.

We have applied global stability analysis, which examines where the traffic system will move to from any initial condition. Much of the analysis of the existence and stability of hypercongested equilibria has instead applied local stability analysis, which examines whether a traffic system that starts in an equilibrium will return to the same equilibrium after a small perturbation. Since the perturbation results in the starting point of the transient dynamics being close to an equilibrium, global analysis is more general. Thus, in the language of the local stability analysis that has been employed



Note: E_1 corresponds to E_1 in Figures 6 and 7, which is a saturated, hypercongested equilibrium. E'_1 is the corresponding equilibrium when there is a moderate increase in demand. E_2 corresponds to E_2 in Figures 6 and 7.

Figure 8: The effects of an increase in demand when the initial steady-state equilibrium is saturated and hypercongested

in much of the debate, the equilibria E_1 and E_3 are stable with both quantity and price perturbations.

We have applied our analysis to examine the stability of steady-state equilibria. But it can also be applied to examine the “comparative statics” of steady states, as well as the path of adjustment between them. Return to Figure 7. Suppose that the traffic system is in steady-state equilibrium at E_1 , and consider the effect of a moderate, once-and-for-all increase in travel demand. This results in a downward shift of the $\dot{T} = 0$ locus, causing the corresponding equilibrium to relocate to a position on the $\dot{C} = 0$ locus southeast of E_1 – call it E'_1 , for which T is higher and C lower. Since E_1 then lies in the interior of area x_1 to the left of the locus MdE_2g , the system moves directly from E_1 to E'_1 (Figure 8 displays the same result in throughput-

trip-price space. Demand increases from D to D' , which results in the saturated equilibrium moving from E_1 to E'_1). Parking remains saturated so that throughput remains unchanged. This requires that t increase, which requires that effective density increase. Since flow equals throughput times $(C+T)/T$, and since $(C+T)/T$ falls, flow decreases. Thus, the increase in demand results in reduced velocity and flow (so that velocity and flow move in the same direction, another indication of hypercongestion) and no change in throughput. Now consider the effect of a large, once-and-for-all increase in travel demand, that causes the $\dot{T} = 0$ locus to move downward so far that no portion lies in the $T - C$ plane. Since the point E_1 is then located in the region x_1 to the right of the locus MdE_2g , the system moves from E_1 to the gridlock equilibrium (see the movement from D to D'' in Figure 8).

6 Discussion

We now attempt to explain in what ways our results differ from those of Verhoef (1999, 2001, and 2003), Small and Chu (2003), and Small and Verhoef (2007), and why. One source of difference is difference in models, which reflect differences in travel contexts. In the bulk of their work, Small and Verhoef had in mind highway or freeway traffic flow, in which dissipative activity takes the form either a vertical queue or a horizontal queue that does not interfere with traffic flow, whether at entry points to the road or at bottlenecks along it. In contrast, we considered downtown traffic in which queuing, in particular cruising for parking, interferes with traffic flow. We do not dispute Small and Verhoef's logic but, even in the context of freeway traffic flow, queues sometimes back up from a bottleneck to the next bottleneck upstream, reducing the latter's capacity; the turbulence generated at entry points reduces the capacity of the freeway (which is the main reason that ramp metering is effective); and when two lanes merge, the efficiency of the merge falls as the time delay associated with it increases, as drivers become increasingly frustrated and aggressive.

Another source of difference between our results and Small and Verhoef's lies in the definition of a hypercongested equilibrium. The difference in definitions stems from our distinction between flow and throughput, which is not present in their models since they do not have cruising for parking. In their models, hypercongestion is present in a steady-state equilibrium if and only if the equilibrium is on the backward-

bending portion of the user cost curve. Thus, they naturally refer to steady-state equilibria on the backward-bending portion of user cost curve as hypercongested equilibria. We define hypercongestion to occur if travel speed is lower than the speed associated with maximum throughput, and a steady-state hypercongested equilibrium to be one in which there is hypercongestion according to this definition. Return to Figure 5. According to our definition, equilibrium E_1 may be either congested or hypercongested, depending on whether travel speed is higher or lower than that associated with maximum throughput.

Based on the assumption that dissipative activity takes the form of a queue that does not interfere with traffic flow, they argue that the backward-bending portion of the user cost curve should be replaced by a vertical segment at capacity flow. While we conducted our analysis without reference to the user cost curve, we recast it in that space. Our user cost curve plotted throughput against user cost, and had four different sections. The first is the upward-sloping section of the user cost curve up to parking capacity; an intersection of the steady-state demand curve with this section of the user cost curve corresponds to a stable, congested equilibrium (it would be E_4 when the parking capacity constraint does not bind). The second is the parking capacity constraint; an intersection of the steady-state demand curve and the parking capacity constraint corresponds to the saturated equilibrium in our analysis. The third is the backward-bending section of the user cost curve for finite trip price; an intersection of the steady-state demand curve and this section of the user cost curve corresponds to the saddlepoint equilibrium E_2 in our analysis.²⁰ The fourth is the section of the user cost curve where user cost is infinite; and intersection of the steady-state demand curve and this section of the user cost curve corresponds to the gridlock equilibrium²¹, E_4 .

Yet another source of difference between our results and Small and Verhoef’s lies in the notion of stability. They conducted local stability analysis, while our stability

²⁰With our choice of functional forms and parameters, the demand curve intersects this section of the user cost curve only once. In general, the demand curve can intersect this section multiple times. We conjecture but have not proved that each of the corresponding intersection points corresponds to an equilibrium that is saddle-path stable, which is consistent with Else’s and Nash’s analyses.

²¹Because we assumed demand to be isoelastic, the demand and user cost curves “intersect” at zero throughput and infinite user cost. Suppose instead that the maximum willingness to pay for travel is finite, so that the demand curve intersects the user cost axis at a finite price. The demand is zero above this price. Thus, once again, the demand and user cost curves intersect at zero throughput and infinite user cost.

analysis was global. Our analysis is more general.

These sources of difference notwithstanding, there remains a more basic point of disagreement between Small and Verhoef on one hand, and ourselves on the other, that the current state of the literature, including our paper, does not resolve. They appear to believe that, at the aggregate level, traffic systems do not respond to an increase in demand by providing reduced throughput, or at least if they do that these situations are practically unimportant. Though our model does not provide a strong case that they are mistaken²², we disagree.²³ Many physical systems respond to increased load with decreased throughput: electrical networks respond to high load with brownouts and blackouts; before fiber-optic technology, long-distance telephone switches used to get jammed with high demand; the absorptive capacity of the environment may fall as the level of pollution increases; etc. We see no reason why traffic systems should not behave similarly. Some recent work provides some support for our view. May, Shepherd, and Bates (2000) used the traffic microsimulation model NEMIS to simulate the effects of an increase in demand on average network speed ($veh - km/veh - hr$) and average network travel ($veh - km/hr$) in hypothetical grid and ring radial networks. For both network configurations, they found that above critical levels of demand both average network speed and average network travel decline – the network analog of travel on the backward-bending portion of the user cost curve. While the paper provided little explanation of this result, we suspect that queue spillbacks and the re-

²²At the end of the previous section, we established that in our model, starting from the well-behaved steady-state equilibrium, E_1 , a sufficiently large once-and-for-all increase in demand leads to gridlock. But since we do not observe gridlock, this theoretical counterexample is hardly compelling.

²³Arnott’s views have changed over the years according to the models he was working on and the contemporaneous literature in transportation engineering journals. When working with André de Palma and Robin Lindsey on the bottleneck model, where a road segment’s capacity is determined by the discharge rate of the segment’s tightest bottleneck and not by the number of cars in the bottleneck queues, he was inclined to the view that hypercongestion is a localized and transient phenomenon occurring within the bottleneck queues. This view was supported by careful analysis of detailed traffic flow data (e.g., Hall, Allen, and Gunter, 1986; Daganzo, Cassidy, and Bertini, 1999) by transportation scientists, and is essentially the same as Small and Verhoef’s current views. His views have changed largely as a result of thinking about downtown traffic congestion, where intersection capacity falls when demand is high due to spillbacks and the increased aggressiveness of drivers in heavily congested traffic, and where cruising for parking and double parking severely interfere with traffic flow. In that traffic context, the larger are the queues at intersections and the “quasi-queues” associated with searching for parking, the lower is system throughput. Looking at highways from that perspective, he has come to the view that similar phenomena occur in highway travel in very heavily congested conditions. The transportation science literature has been undergoing a similar change in perspective (e.g., Varaiya, 2008, on ramp metering; Daganzo, 2002, on merging traffic streams downstream of on-ramps; and Lo and Szeto, 2005, on spillbacks).

duced capacity of intersections with increased demand are responsible. Lo and Szeto (2005) demonstrates in a dynamic, physical queuing network model how spillbacks can generate hypercongestion. Since both papers employ models that are dynamic in Verhoef’s terminology, they do not however establish the existence of stable, hypercongested, steady-state equilibria.

Where does all this leave us in terms of the debate on the possibility of stable, hypercongested, steady-state traffic equilibria? This paper has contributed to the debate by providing a thorough analysis of a quite particular model. It has not resolved the debate, and indeed we are not sure that the debate will ever be resolved since the terms of the debate may keep changing as our mathematical tools and understanding of traffic flow become more sophisticated. The debate has nonetheless been fruitful since it has demonstrated the importance of precision in definition and analysis in this context and has improved our understanding of both the economics and physics of traffic flow.

A Appendix

A.1 Trajectories

This appendix presents some trajectories of the differential equation system. For expositional convenience, we shall make a transformation of variables and reduce the 3D system to a 2D system. The proper transformation is defined as follows. Define

$$R = R^+ + R^- \quad , \quad (\text{A-1})$$

where

$$R^+ = \max\{R, 0\} = \frac{R + |R|}{2} \quad (\text{A-2})$$

$$R^- = \min\{R, 0\} = \frac{R - |R|}{2} \quad . \quad (\text{A-3})$$

Let $C(u) = R^+(u)$ and $S(u) = P + R^-(u)$. Note that when $R(u) \geq 0$, $C(u) = R(u)$ and $S(u) = P$, and when $R(u) \leq 0$, $C(u) = 0$ and $S(u) = P + R(u)$. The transformed

autonomous differential equation system is given by

$$\dot{T}(u) = D(\rho(mt(T(u), \frac{1}{2}(|R| + R), P) + \frac{\frac{1}{2}(|R| + R)l}{P}) + \lambda l) - \frac{T(u)}{mt(T(u), \frac{1}{2}(|R| + R), P)} \quad (\text{A-4})$$

$$\dot{R}(u) = \frac{T(u)}{mt(T(u), \frac{1}{2}(|R| + R), P)} - \frac{P + \frac{1}{2}(R - |R|)}{l} \quad (\text{A-5})$$

Since this system is Leibnitz, all existence and uniqueness theorems apply. However, the system is only piecewise differentiable and there is a phase transition at $R^+(u) = 0$.

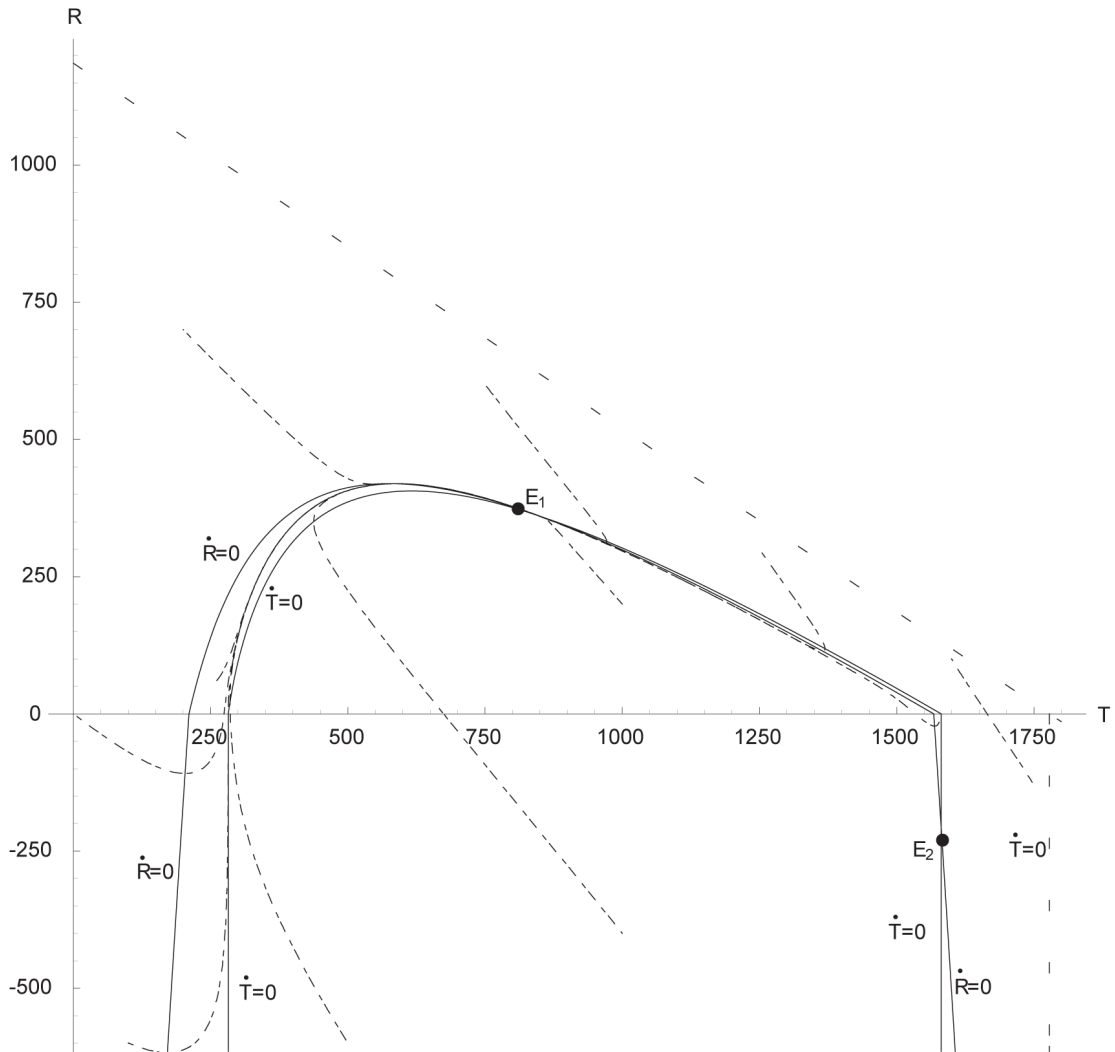


Figure 9: Trajectories of the transformed differential equation system

Geometrically, this transformation corresponds to making the vertical portion of Figure 7 horizontal and stretching C and S accordingly. The transformed system is shown in Figure

9, which assumes the parameter values stated in Section 4.1. In this figure, the $\dot{T}(u) = 0$ and $\dot{R}(u) = 0$ loci are shown with solid lines and trajectories with dotted lines. We do not show E3 since doing so would entail loss of important detail, but one should note that the $\dot{R}(u) = 0$ locus cuts the x -axis when $\{T, R\} = \{1778.17, -3712\}$, which corresponds to jam density. However, we shown one of the trajectories approaching E_3 at the far right of the figure.

A.2 Derivation of $E = T/(mt)$

This appendix derives the equilibrium condition $E = T/(mt)$. Let A (arrivals) denote the cumulative number of cars that have entered the downtown area, X (exits) denote the cumulative number of cars that have exited the downtown area, and S the stock of occupied parking spaces. We have the stock identity that $A = T + C + S + X$. Thus, $\dot{A} = \dot{T} + \dot{C} + \dot{S} + \dot{X}$. Moreover, D is the entry rate into downtown and thus $\dot{A} = D$. Since visit lengths are Poisson distributed, with mean l , we have $\dot{X} = S/l$. Letting E denote the exit rate from the in-transit pool and Z the exit rate from the cruising-for-parking pool, we have that $\dot{T} = D - E$, $\dot{C} = E - Z$, and $\dot{S} = Z - S/l$. We have two régimes to consider. In the saturated parking régime:

$$Z = \frac{P}{l} \quad (\text{A-6})$$

$$\dot{P} = 0 \quad , \quad (\text{A-7})$$

whereas in the unsaturated parking régime:

$$Z = E \quad (\text{A-8})$$

$$\dot{S} = E - \frac{S}{l} \quad . \quad (\text{A-9})$$

From these equations, it is evident that the entire evolution can be determined once E is calculated. Let M be the cumulative number of cars that have exited the in-transit pool, so that $E = \dot{M}$ and $T = A - M$. The technology of traffic congestion is captured by the function $t = t(T, C, P)$ or alternatively by $v = v(T, C, P)$. Consider a car that enters the in-transit pool at time u . By time w it has traveled a distance

$$x(u, w) = \int_u^w v(y) dy \quad . \quad (\text{A-10})$$

Thus, of the cars that enter the in-transit pool at time u , the proportion that have exited it by time w , $\Xi(w, u)$, is

$$\Xi(w, u) = 1 - e^{-hx(u, w)} \quad , \quad (\text{A-11})$$

where $h = 1/m$, and so the number that have exited it by time m , $N(w, u)$, is

$$N(w, u) = D(u)(1 - e^{-hx(u, w)}) \quad . \quad (\text{A-12})$$

As a result, the total number of cars that have exited the in-transit pool by time w , $M(w)$, is

$$M(w) = \int_0^w D(u)(1 - e^{-hx(u,w)})du \quad . \quad (\text{A-13})$$

Differentiating this with respect to w , we obtain

$$\dot{M}(w) = D(w)(1 - e^{-hx(w,w)}) + \int_0^w D(u)hx_w(u, w)e^{-hx(u,w)}du \quad . \quad (\text{A-14})$$

Since the first term on the right hand side is zero, we have

$$\dot{M}(w) = \int_0^w D(u)hx_w(u, w)e^{-hx(u,w)}du \quad . \quad (\text{A-15})$$

From (A-11), $x_w(u, w) = v(w)$. This comes out of the integral, so that we have

$$\dot{M}(w) = v(w) \int_u^w D(u)he^{-hx(u,w)}du \quad . \quad (\text{A-16})$$

The total number of cars that have entered the in-transit pool by time w is simply

$$A(w) = \int_0^w D(u)du \quad . \quad (\text{A-17})$$

Thus, the stock of cars in the in-transit pool at time w is

$$T(w) = \int_0^w D(u)e^{-hx(u,w)}du \quad . \quad (\text{A-18})$$

Combining this with (A-16), we obtain

$$\dot{M}(w) = v(w)hT(w) \quad , \quad (\text{A-19})$$

and thus $E = vhT$ or

$$E = \frac{T}{mt} \quad . \quad (\text{A-20})$$

A.3 Equilibrium and stability with no parking constraint

The model that has been the focus of the debate over the existence and stability of steady-state hypercongested equilibria does not contain parking. It is therefore natural to enquire into the existence and stability of hypercongested equilibria in a simplified version of our

model without parking. This is a special case of our model for which the parking constraint does not bind, the traffic system is always in régime 2, and the parking fee is zero. Modifying the example so that the demand function here with $\lambda = 0$ is the same as before with $\lambda = 1$, Figures 4 and 5 continue to apply, except that the parking capacity constraint does not bind. There are then three equilibrium E_4 , E_3 , and E_2 . E_4 is locally stable and is reached starting from any initial point to the left of the locus dE_2g in Figure 4; E_3 is the locally stable, gridlock equilibrium, and is reached from any initial point to the right of the locus dE_2g ; and E_2 is the saddle-path stable equilibrium, and is reached from any initial point on the locus dE_2g . The issue of central interest is whether traffic in the equilibria E_4 and E_2 is congested or hypercongested. Since there are no cars cruising for parking, flow and throughput coincide, and traffic is congested when $T < V_j/2$ and is hypercongested when the inequality is reversed. Since $T = V_j/2$ at the peak of the $\dot{S} = 0$ locus, as Figure 4 is drawn E_4 is congested and E_2 is hypercongested. But from Figure 5, it can be seen that if demand is increased so that E_4 lies on the backward-bending portion of the user cost curve, all the equilibria are hypercongested. And above a critical level of demand, the equilibria E_2 and E_4 disappear and only the gridlock equilibrium E_3 remains.

References

- [1] Agnew, C. 1977. The theory of congestion tolls. *Journal of Regional Science* 17: 381-393.
- [2] Arnott, R. 2006. Spatial competition between parking garages and downtown parking policy. *Transport Policy* 13: 458-469.
- [3] Arnott, R. and E. Inci. 2006. An integrated model of downtown parking and traffic congestion. *Journal of Urban Economics* 60: 418-442.
- [4] Arnott, R. and J. Rowse. 1999. Modeling parking. *Journal of Urban Economics* 45: 97-124.
- [5] Arnott, R. and J. Rowse. forthcoming. Downtown parking in auto city. *Regional Science and Urban Economics*.
- [6] Chu, X. 1995. Endogenous trip scheduling: The Henderson approach reformulated and compared with the Vickrey approach. *Journal of Urban Economics* 37: 324-343.
- [7] Daganzo, C. 1992. The cell transmission model. Part I: A simplified representation of highway traffic. *California Partners for Advanced Transit and Highways (PATH). Research Reports: Paper UCB-ITS-PRR-93-7*.
- [8] Daganzo, C. 2002. A behavioral theory of multi-lane traffic flow. Part II: Merges and the onset of congestion. *Transportation Research B* 36: 159-169.
- [9] Daganzo, C., M. Cassidy and R. Bertini. 1999. Possible explanations of phase transitions in highway traffic. *Transportation Research A* 33: 365-379.

- [10] Dewees, D. 1978. Estimating the time costs of highway congestion. *Econometrica* 47: 1499-1512.
- [11] Else, P. 1981. A reformulation of the theory of optimal congestion taxes. *Journal of Transport Economics and Policy* 15: 217-232.
- [12] Else, P. 1982. A reformulation of the theory of optimal congestion taxes: a rejoinder. *Journal of Transport Economics and Policy* 16: 299-304.
- [13] Greenshield, B. 1935. A study of traffic capacity. *Highway Research Board Proceedings* 14: 448-477.
- [14] Hall, F., B. Allen and M. Gunter. 1986. Empirical analysis of freeway flow-density relationships. *Transportation Research A* 20: 197-210.
- [15] Henderson, J. 1981. The economics of staggered work hours. *Journal of Urban Economics* 9: 349-364.
- [16] Johnson, M. 1964. On the economics of road congestion. *Econometrica* 32: 137-150.
- [17] Lindsey, R. 1980. Non-steady-state traffic flow. Directed research project in partial fulfillment of the requirements for the Degree of PhD, Department of Economics, Princeton University.
- [18] Lo, H. and W. Szeto. 2005. Road pricing modeling for hyper-congestion. *Transportation Research A* 39: 705-722.
- [19] May, A., S. Shepherd, and J. Bates. 2000. Supply curves for urban road networks. *Journal of Transport Economics and Policy* 34: 261-290.
- [20] McDonald, J. and E. d’Ouille. 1988. Highway traffic flow and the ‘uneconomic’ region of production. *Regional Science and Urban Economics* 18: 503-509.
- [21] Nash, C. 1982. A reformulation of the theory of optimal congestion taxes: a comment. *Journal of Transport Economics and Policy* 26, 295-299.
- [22] Neuberger, H. 1971. The economics of heavily congested roads. *Transportation Research* 5: 283-293.
- [23] Small, K. and E. Verhoef. 2007. *The Economics of Urban Transportation*. Abingdon, England: Routledge.
- [24] Small, K. and X. Chu. 2003. Hypercongestion. *Journal of Transport Economics and Policy* 37: 319-352.
- [25] Varaiya, P. 2008. Congestion, ramp metering and tolls. *Philosophical transactions of the Royal Society* 366: 1921-1930.
- [26] Verhoef, E. 1999. Time, speeds, flows and densities in static models of road traffic congestion and congestion pricing. *Regional Science and Urban Economics* 29: 341-369.

- [27] Verhoef, E. 2001. An integrated dynamic model of road traffic congestion based on simple car-following theory: exploring hypercongestion. *Journal of Urban Economics* 49: 505-542.
- [28] Verhoef, E. 2003. Inside the queue: hypercongestion and road pricing in a continuous time-continuous place model of traffic congestion. *Journal of Urban Economics* 54: 531-565.
- [29] Vickrey, W. 1969. Congestion theory and transport investment. *American Economic Review* 59: 251-261.
- [30] Walters, A. 1961. The theory and measurement of private and social cost of highway congestion. *Econometrica* 29: 676-699.