WORKING PAPER

ITLS-WP-17-05(1)

Continuous-time general link transmission model with simplified fanning,

Part I: Theory and link model formulation

By
Michiel C.J. Bliemer and Mark P.H. Raadsen

Institute of Transport and Logistics Studies (ITLS),
The University of Sydney Business School, Sydney, Australia

March 2017

ISSN 1832-570X

INSTITUTE of TRANSPORT and LOGISTICS STUDIES

The Australian Key Centre in Transport and Logistics Management

The University of Sydney

Established under the Australian Research Council's Key Centre Program.

NUMBER: Working Paper ITLS-WP-17-05(1)

TITLE:

Continuous-time general link transmission model with simplified fanning,

Part I: Theory and link model formulation

ABSTRACT:

The kinematic wave theory is widely used to simulate traffic flows on road segments. Link transmission models are methods to find a solution to the kinematic wave model, however, their computational efficiency heavily relies on the shape of the fundamental diagram that is used as input. Despite the limitations and drawbacks of triangular and piecewise linear fundamental diagrams, they remain popular because they result in highly efficient algorithms. Using smooth nonlinear branches is often preferred in terms of realism and other desirable properties, but this comes at a significantly higher computational cost and requires time discretisation to find an approximate solution. In this paper we consider a nonlinear fundamental diagram as input and propose on-the-fly multi-step linearization techniques to simplify expansion fans. This leads to two simplified link transmission models that can be solved exactly in continuous time under the assumption of piecewise stationary travel demand. One of the models simplifies to shockwave theory in case of a single step. We show that embedding shockwave theory in the link transmission model allows for finding an exact solution in continuous time and we discuss the potential for the design of efficient event-based algorithms for general networks.

KEY WORDS: ***Kinematic wave theory, link transmission model, nonlinear concave fundamental diagram, on-the-fly linearization, simplified fanning, shockwave theory, event-based algorithm***

AUTHORS: **Bliemer and Raadsen**

Acknowledgements: This research was partially funded by the Australian Research Council, Linkage Project grant LP LP130101048.

CONTACT:

INSTITUTE OF TRANSPORT AND LOGISTICS STUDIES (H73)

The Australian Key Centre in Transport and Logistics Management

The University of Sydney NSW 2006 Australia

Telephone: +612 9114 1824

E-mail: business.itlsinfo@sydney.edu.au

Internet: http://sydney.edu.au/business/itls

DATE: March 2017

1. Introduction

1.1 Dynamic traffic assignment framework

Dynamic traffic assignment (DTA) models are used as tools to provide forecasts of traffic conditions for traffic management and are increasingly applied in strategic transport planning to support decision making regarding infrastructure investments. The advantage of dynamic models over static models is that they are capable of capturing time-varying travel demand and describe important traffic flow dynamics such as queue build up, dissipation and spillback. Unfortunately, dynamic models have a much higher computational complexity, mainly due to the more advanced network loading applied in such (mostly simulation based) models. A large body of research exists that looks into increasing the efficiency of simulating traffic on large transportation networks while retaining the most essential characteristics of traffic flow.

Different representations of traffic flow exist, namely microscopic, mesosopic and macroscopic. In this paper we focus on macroscopic models. A typical macroscopic simulation-based DTA framework consists of a route choice (and possibly departure time choice) component that yields path flows for each origin-destination pair, and a dynamic network loading (DNL) component that simulates these path flows on a given transport network consisting of links and nodes and yields path travel times. Each link represents a homogeneous spatial road segment and has an associated fundamental diagram (FD) that describes macroscopic traffic flow behaviour. Nodes are represented as non-spatial elements that connect the boundaries of multiple links and may have associated traffic signal controls.

A DNL model consists of two components, namely a link model and a node model that describe the traffic flow interactions on the link level and node level, respectively. The link model determines how vehicles and traffic states propagate across the link, impacting sending flows (available demand for link outflows) and receiving flows (available supply for link inflows), while the node model considers all sending and receiving flows and determines actual link inflows and outflows.

1.2 First order dynamic network loading models

A popular first order link model is the hydrodynamic kinematic wave model (Lighthill and Whitham, 1955; Richards, 1956). Several methods have been described in the literature that solve the kinematic wave model to simulate traffic and that can be applied to transport networks. The two most well-known methods are the *cell transmission model* (CTM) as proposed by Daganzo (1994, 1995) and the *link transmission model* (LTM) proposed by Yperman et al. (2005) using theories presented by Newell (1993a,b,c). The word ‘model’ is somewhat ambiguous and confusing, since CTM and LTM can be seen as two different solution schemes or algorithms that essentially solve the same first order DNL problem as described by a combination of the kinematic wave model and a node model.

There are several differences between CTM and LTM with respect to solution approach and assumptions on the FD. CTM considers the partial differential equation formulation of the kinematic wave model and uses the first order finite Godunov scheme as solution approach (Lebacque, 1996). As a result, CTM describes traffic in terms of average densities over spatial cells across the link and propagates vehicles according to constraints on flow and storage capacity of each cell. On the other hand, LTM considers the Hamilton-Jacobi equation formulation and (implicitly or explicitly) uses the Lax-Hopf formula to find a solution and represents traffic in terms of (cumulative) flows only at the link boundaries in which traffic states are propagated upstream and downstream.

These methodological differences between CTM and LTM have important consequences with respect to model efficiency, accuracy, flexibility, and applicability. CTM requires discretising space into cells, which makes this approach less efficient and less accurate than LTM. On the other hand, CTM allows more flexibility with respect to the shape of the FD and the node model, while LTM requires a concave FD and a node model that satisfies the invariance principle (see Lebacque and Khoshyaran, 2005). In this paper we consider the LTM approach as an efficient method for DNL on large networks.

1.3 Literature review on link transmission models

We can distinguish LTM formulations by the assumed shape of the FD. An FD describes flow (also referred to as flux) as a function of density and consists of a hypocritical branch and a hypercritical branch (see e.g., Cascetta, 2009). The hypocritical branch represents (steady-state) uncongested traffic states in which flow is constrained by demand and hence flow increases with density. The hypercritical branch represents (steady-state) congested traffic states in which flow is constrained by supply and hence flow decreases with density. LTM requires that both branches of the FD are concave.

The hypocritical or hypercritical branch are typically one of the following, (i) linear, (ii) piecewise linear, or (iii) smooth nonlinear. Piecewise linear branches are not differentiable everywhere and hence are considered non-smooth, as opposed to smooth nonlinear branches that are continuously differentiable.

Based on the work of Newell (1993a), Yperman et al. (2005) proposed LTM assuming linear branches for both the hypocritical as well as the hypercritical branch (resulting in a triangular FD) and used a fixed time discretisation scheme to solve the model. A linear hypocritical branch means that the speed of a vehicle is fixed to the maximum speed for all hypocritical states, which is not realistic. Nevertheless, the assumption of linear branches has been adopted by many others (i.e., in both LTM and CTM) mainly for computational convenience. Yperman (2007) proposed an extension to include piecewise linear branches, which allows for multiple kinematic wave speeds and vehicle speeds while only slightly increasing computational complexity. The equations provided do actually not yield a physically realistic solution (see e.g., Bliemer et al., 2016), but this can easily be amended.

The algorithm of Yperman et al. (2005) requires relatively small discrete time steps that do not exceed the free-flow travel times of the links in order to satisfy the Courant-Friedrichs-Lewy condition. To increase the computational efficiency, Himpe et al. (2016) describe an iterative algorithm for LTM that does not impose such a hard upper limit on the time step size. Both algorithms rely on a time grid and therefore yield approximate solutions (although arguably more accurately than CTM).

Gentile (2010) extended the model formulation of Yperman et al. (2005) to smooth nonlinear branches and called it *general LTM*. While non-linear branches are considered to be more able to describe actual traffic operations (particularly in the hypocritical branch), LTM with non-linear branches is mathematically and computationally more complex, mainly due to fanning effects in case of increasing flows. As a result, only few others seem to have adopted non-linear branches in applications of LTM.

Since the kinematic wave model may have multiple solutions since the placement of shocks may not be unique, Luke (1972) and Newell (1993a) proposed a minimum principle in order to find the ‘correct’ or physically most meaningful solution, which was also adopted in LTM. The physically most meaningful solution is typically considered the one that satisfies the entropy condition. Such entropic solutions are unique, and Daganzo (2005a) provides a formal proof using variational theory that the Luke-Newell minimum principle indeed results in such an entropic solution. In a more mathematical stream of research, Aubin et al. (2008) and Claudel and Bayen (2010a,b) use viability theory and the Lax-Hopf formula to further generalise the solution to the Hamilton-Jacobi equations by also including initial and internal boundary conditions. Friesz et al. (2013) discuss the Lax-Hopf solution to different formulations of the kinematic wave model assuming a general concave FD.

While LTM is traditionally solved assuming some time discretisation, exact entropic solutions to the kinematic wave model have been proposed under the assumption of piecewise affine initial and boundary conditions. There exist event based front tracking algorithms (e.g., Wong and Wong, 2002; Lu et al., 2008) that work with concave and non-concave FDs. These event based algorithms track changes in traffic states along a link. There also exist non-event based grid-free methods such as proposed by Mazaré et al. (2011) who adopt the Lax-Hopf formula as an efficient computational method for finding an exact solution for a specific time instant and location given a concave FD. It should be

noted that these methods focus on a single link and may not necessarily translate to efficient algorithms for general networks.

Only recently it was formally shown by Jin (2015) and Han et al. (2016b) that (discrete time) LTM is essentially a special case of the (continuous time) Lax-Hopf solution to the kinematic wave model. While there have been continuous-time formulations for solving the kinematic wave model based on a general concave FD (see e g., Han et al., 2016a), the connection with the general LTM formulation in Gentile (2010) is not directly obvious.

In order to find an exact solution on an entire network, Raadsen et al. (2016) considered continuous-time LTM with a triangular FD and proposed an event-based algorithm adapting ideas from front tracking and other grid-free solutions. This algorithm does not track changes in traffic states along a link over time, but rather only calculates traffic state changes at the link boundaries over time. It is a relatively simple algorithm that, for each traffic state change at the upstream or downstream end of the link, predicts the time instant at which this traffic state change would occur on the opposite end of the link. Using a sorted event list, events are checked in chronological order, and if an event is validated the corresponding traffic state is sent to the node model and resulting flow rates are updated for all relevant links. If the event is not valid (because another event happened prior to the predicted time instant), it is either updated or discarded. In order to efficiently apply the algorithm on large networks, they put a flow change acceptance threshold in place that does not generate an event if the flow rate change is below a certain threshold (e.g., 1 veh/h). Due to the simplicity of the calculations, the algorithm is capable of generating and processing around 100,000 events per second on a regular desktop computer, such that DNL based on split proportions on large scale networks is possible within one minute.

1.4 Assumptions and aim of the study

In this paper we make the following assumptions: (i) only a single vehicle type is considered, (ii) each link is described by a two-regime concave FD with smooth nonlinear branches, (iii) the node model satisfies all first order conditions as outlined in Tampère et al. (2011) and does not consider traffic controls, and (iv) the travel demand is temporarily stationary (i.e., piecewise affine boundary conditions).

The aim of our study is to establish a continuous-time LTM formulation considering an FD with smooth nonlinear branches that allows determining an exact solution to the kinematic wave model, and that allows the development of efficient algorithms on general networks. In other words, we aim to extend the event-based algorithm proposed in Raadsen et al. (2016) that was specifically developed for the triangular FD to include nonlinear FDs. Such an extension is not straightforward since in many cases an infinite number of wave speeds will occur when modelling a flow state transition instead of just the one wave speed in case of a triangular FD. In other words, we must consider the occurrence of expansion fans in entropic solutions when flow increases. These expansion fans make finding an exact solution in a network over a period of time challenging.

In this paper we approach this challenge by looking for an exact non-entropic solution to the kinematic wave model. As argued in Henn (2003), such solutions may be easier to calculate, although there will be some loss in physical relevance. Two non-entropic solutions in particular are of interest, namely solutions that are consistent with shockwave theory (also referred to as shockwave analysis, see e.g., May, 1990), and solutions that are based on a scheme as proposed in the front tracking algorithm in Henn (2003). Both approaches only simplify expansion fans, such that the solution is in fact entropic when flows do not increase, and is approximate entropic when flows increase. As will be shown in this paper, both non-entropic solutions result from a specific approximation (linearization) of the FD. As stated by Daganzo (2001), approximation of FDs introduce an error that does not increase in space or time, and hence these solutions are expected to be stable. Therefore, when making the choice between finding an entropic solution assuming a simplified (triangular or piecewise linear) FD or finding a non-entropic solutions as discussed above assuming a non-simplified (nonlinear) FD, we believe that the

latter should be preferred since it only affects (cumulative) flows temporarily during the processing of simplified expansion fans, while the former simplifies and affects all (cumulative) flows throughout the simulation.

1.5 Contributions and paper outline

Our contributions are as follows. First, we argue why using a smooth nonlinear hypocritical branch is important, and we establish a simple dual-quadratic FD with five physically meaningful parameters. Secondly, we derive the general continuous-time LTM equations from the Lax-Hopf formula to explicitly show the relationship, and to use as a starting point for our discussion. Thirdly, we propose two methods for on-the-fly linearization of the FD, which allows us to simplify expansion fans and find an exact solution. One of these models can be seen as an extension of shockwave theory, and the other is consistent with the scheme proposed in Henn (2003). Finally, we discuss the potential of using event-based algorithms to solve LTM-based DNL for general networks, and we argue that solving for a non-entropic solution considering nonlinear FDs using shockwave theory is computationally more efficient than determining an entropic solution considering piecewise linear FDs. This paves the way to develop efficient event-based algorithms similar to, but more general than, the one proposed by Raadsen et al. (2016) for triangular FDs. The focus of this paper is on the link model formulation and theory (which also provides theory behind the algorithm presented in Raadsen et al., 2016), while in the Part II companion paper Raadsen and Bliemer (2017) an event-based algorithm for networks is developed, within-iteration route choice methodology is added and numerical results are shown.

The paper is outlined as follows. Section 2 discusses the importance of the shape of the FD and proposes a simple two-regime nonlinear FD. In Section 3 we discuss a general LTM framework and introduce the relevant variables. In Section 4 we derive the main continuous-time equations of the link model from the Lax-Hopf formula. In order to be able to solve the model exactly, we assume piecewise stationary travel demand in Section 5 and propose two linearization methods to simplify expansion fans that guarantee that all cumulative inflow and outflow curves are piecewise linear. In Section 6 we rewrite the link model to a lower envelope problem that can be efficiently and exactly solved by sequentially determining the next shockwave that arrives at the opposite border. Finally, Section 7 discusses computational efficiency under the two linearization methods and compares this with the efficiency when using linear and piecewise linear FDs.

2. Shape of the fundamental diagram

2.1 Considerations regarding the shape of the fundamental diagram

The assumed shape of the FD has important implications regarding a number of properties in DTA, namely (i) computational efficiency, (ii) realism in terms of the traffic flows, speeds, and travel times, (iii) uniqueness of a traffic equilibrium, and (iv) ease of calibration of the parameters.

Here we discuss different shapes of the FD based on the properties mentioned above. Figure 1 shows several popular shapes described by function $q = \Phi(k)$ where q is flow (veh/h) and k is density (veh/km). All shapes can be accommodated by CTM. The Lax-Hopf formula requires a concave FD (see also Appendix A), hence the shape in Figure 2(a) cannot be considered in LTM.

First, we consider *computational efficiency*. When using LTM to solve the kinematic wave model, single-regime FDs as shown in Figures 1(d)–(e) are less computationally efficient than two-regime FDs. This is due to the fact that in single-regime diagrams the forward and backward kinematic wave speeds approach zero as the flow reaches capacity. This means that the algorithm needs to keep track of, and potentially search through, the entire history of link inflows and outflows in order to determine the flow rates in the current time step or time instant, as illustrated by the equations in Gentile (2010). Further, FDs with linear or piecewise linear branches, such as depicted in Figures 1(a)–(b) and 1(f)–(g), result

in more computationally efficient algorithms than using FDs with nonlinear branches due the a limited number of traffic states represented in expansion fans.

Next, we look at *realism*. A linear hypocritical branch as present in the triangular and trapezoidal FDs in Figures 1(b) and 1(c), respectively, is often criticized for the fact that speed is constant over the entire range of hypocritical flows. One would expect for example that for motorways with a maximum speed around 120 km/h, the speed will gradually decline with increasing flows until the speed has reached a critical level of about 80 km/h. On the other hand, a linear hypercritical branch is typically considered not as problematic, and some will even argue that this branch should rather be convex than concave. Therefore, a linear hypercritical branch is often considered a reasonable assumption. A symmetric diagram as seen in Greenshields' FD shown in Figure 1(d) is also considered unrealistic, as one would expect an asymmetric shape due to the fact that forward kinematic waves move much faster than backward kinematic waves. Finally, a piecewise linear hypocritical branch results in a relationship between travel time and flow that is non-convex, which seems counter-intuitive.

With respect to *uniqueness of a deterministic user equilibrium* in general transport networks, we note that a necessary (but not sufficient) condition is strictly increasing travel times with flow. FDs with a strictly concave hypocritical branch satisfy this condition, i.e. diagrams as shown in Figures 1(a), 1(d)–(e), 1(h)–(i). FDs with a linear or piecewise linear hypocritical branch (i.e., diagrams shown in Figures 1(b)–(c), and 1(f)–(g)) do not satisfy this condition. While this is considered an important property in static traffic assignment, in dynamic assignment this issue is – perhaps surprisingly – largely ignored. A pragmatic approach proposed by Boyles et al. (2013) that allows a unique user equilibrium in most cases is to create a piecewise linear hypocritical branch with a very short segment near zero flow.

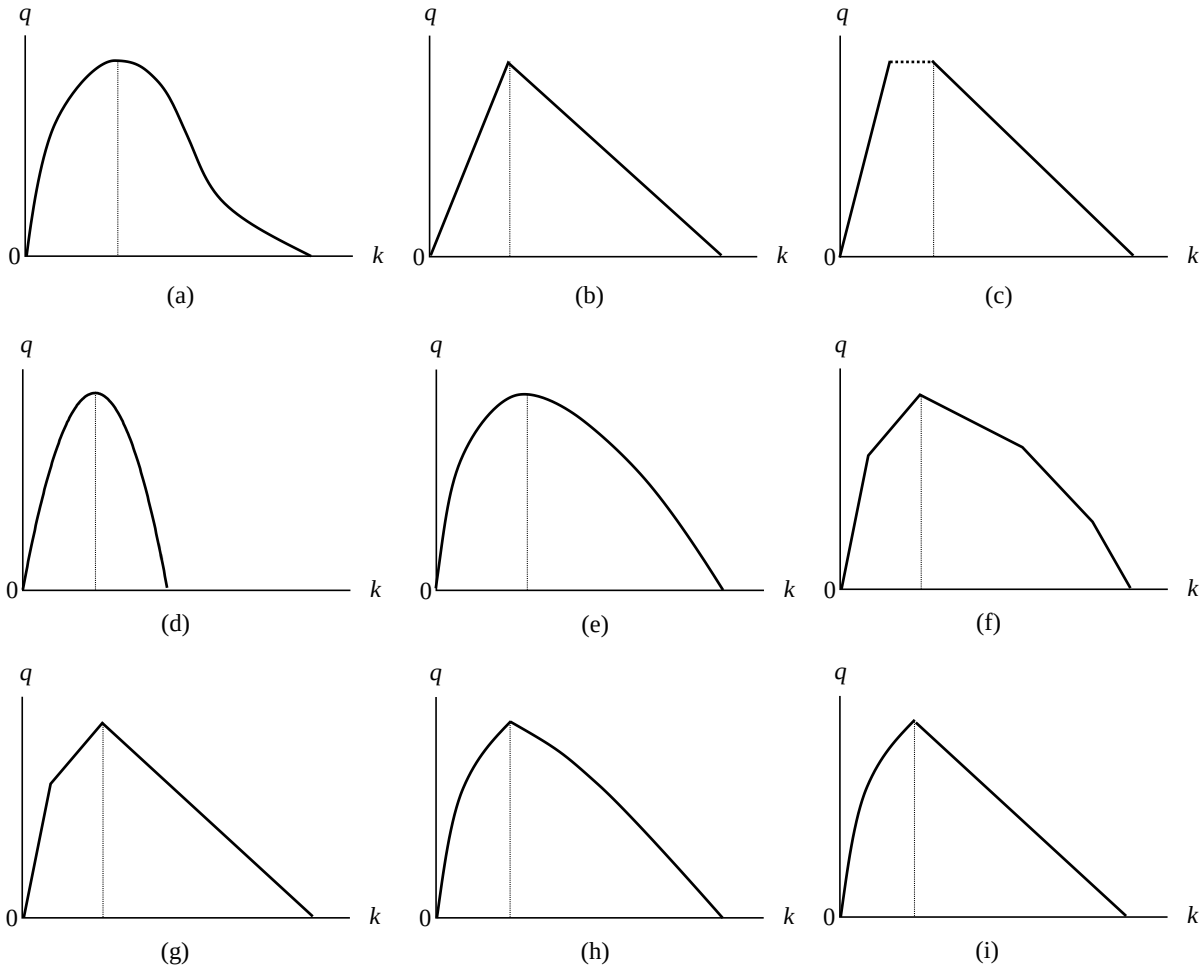


Figure 1: Different shapes of fundamental diagram

Lastly, we look at *ease of calibration of the parameters* in the flux function, which is important for large scale networks. While piecewise linear branches provide a lot of flexibility in the shape of the FD, it comes at a cost of a large number of parameters that are not that easy to determine (and are often set in a rather arbitrary fashion). In contrast, a non-linear differentiable hypocritical branch can easily be specified based on only the maximum speed, critical speed, and capacity. Further, FDs that have too few degrees of freedom, such as the triangular and Greenshields' diagram, will be easy to specify but will be more difficult to fit to empirical data (and hence lack the realism, as mentioned earlier).

To summarise, we would prefer a smooth nonlinear hypocritical branch for realism, uniqueness of a traffic equilibrium, and ease of calibration, but we prefer (piecewise) linear branches for computational efficiency. In this paper we aim to get the best of both worlds by considering a two-regime nonlinear FD, at least in the hypocritical branch as illustrated in Figures 1(h)–(i), while we simplify expansions fans that will result in a computational efficiency on networks similar to (piecewise) linear FDs.

In the next subsection we propose a simple functional form for an FD of the shape shown in Figure 1(h) and also consider the case where the hypercritical branch is considered linear (depicted in Figure 1(i)).

2.2 Dual-quadratic fundamental diagram

Lu et al. (2008) proposed a piecewise quadratic FD, with as special case a two-regime quadratic FD. Here we explicitly describe such a dual-quadratic FD defined by a two-regime flux function $\Phi(k)$ that has five physically meaningful calibration parameters that are relatively easy to determine, namely (i) maximum wave speed $\gamma^{\max}$ (km/h) (which equals the maximum allowed/desired vehicle speed), (ii) capacity Q (veh/h), (iii) jam density K (veh/km), (iv) critical vehicle speed σ^{crit} (km/h), and (v) minimum wave speed $\gamma^{\min}$ (km/h). Denote the critical density by k^{crit} (veh/km), which is defined as $k^{\text{crit}} = Q / \sigma^{\text{crit}}$.

We define a two-regime flux function as follows:

$$\Phi(k) = \begin{cases} \Phi_I(k), & \text{if } 0 \le k \le k^{\text{crit}}, \\ \Phi_{II}(k), & \text{if } k^{\text{crit}} \le k \le K, \end{cases} \tag{1}$$

where $\Phi_I(k^{\text{crit}}) = \Phi_{II}(k^{\text{crit}})$, and where the hypocritical and hypercritical functions are given functions $\Phi_I : [0, k^{\text{crit}}]$ (veh/km) $\to [0, Q]$ (veh/h) and $\Phi_{II} : [k^{\text{crit}}, K]$ (veh/km) $\to [0, Q]$ (veh/h).

In case of a dual-quadratic FD, these functions are given by

$$\Phi_I(k) = \left(\gamma^{\max} - \alpha k\right)k, \qquad \text{with } \alpha = \frac{\sigma^{\text{crit}}}{Q}\left(\gamma^{\max} - \sigma^{\text{crit}}\right), \tag{2}$$

$$\Phi_{II}(k) = \left(\gamma^{\min} + \beta(k - 2K)\right)k + \left(\beta K - \gamma^{\min}\right)K, \qquad \text{with } \beta = \frac{Q + \gamma^{\min}\left(K - \dfrac{Q}{\sigma^{\text{crit}}}\right)}{\left(K - \dfrac{Q}{\sigma^{\text{crit}}}\right)^2}. \tag{3}$$

The vehicle speeds $\sigma_I : [0, k^{\text{crit}}]$ (veh/km) $\to [\sigma^{\text{crit}}, \gamma^{\max}]$ (km/h) and $\sigma_{II} : [k^{\text{crit}}, K]$ (veh/km) $\to [0, \sigma^{\text{crit}}]$ (km/h) can be computed as

$$\sigma_I(k) = \frac{\Phi_I(k)}{k}, \quad \text{and} \quad \sigma_{II}(k) = \frac{\Phi_{II}(k)}{k}. \tag{4}$$

Denote the maximum vehicle speed by $\sigma^{\max}$ (km/h). By definition it holds that $\sigma^{\max} = \gamma^{\max}$ since due to L'Hôpital's rule it follows that $\sigma^{\max} = \lim_{k \to 0} \sigma_I(k) = \lim_{k \to 0} d\Phi_I(k) / dk = \gamma^{\max}$.

Since boundary conditions in LTM are expressed in terms of flow rates, we also require the inverse flux functions $\Phi_I^{-1}:[0,Q]$ (veh/h) $\rightarrow[0,k^{\text{crit}}]$ (veh/km) and $\Phi_{II}^{-1}:[0,Q]$ (veh/h) $\rightarrow[k^{\text{crit}},K]$ (veh/km), defined by

$$\Phi_I^{-1}(q)=\frac{1}{2\alpha}\left(\gamma^{\max}-\sqrt{4\alpha q+(\gamma^{\max})^2}\right), \tag{5}$$

$$\Phi_{II}^{-1}(q)=K-\frac{1}{2\beta}\left(\gamma^{\min}+\sqrt{4\beta q+(\gamma^{\min})^2}\right). \tag{6}$$

An illustration of the dual-quadratic FD is shown in Figure 2. Note that in order to have a strictly increasing concave hypocritical branch it is required that $1\le\gamma^{\max}/\sigma^{\text{crit}}<2$, and for the hypercritical branch to be concave and strictly decreasing it is required that $1\le\gamma^{\min}(Q/\sigma^{\text{crit}}-K)/Q<2$.

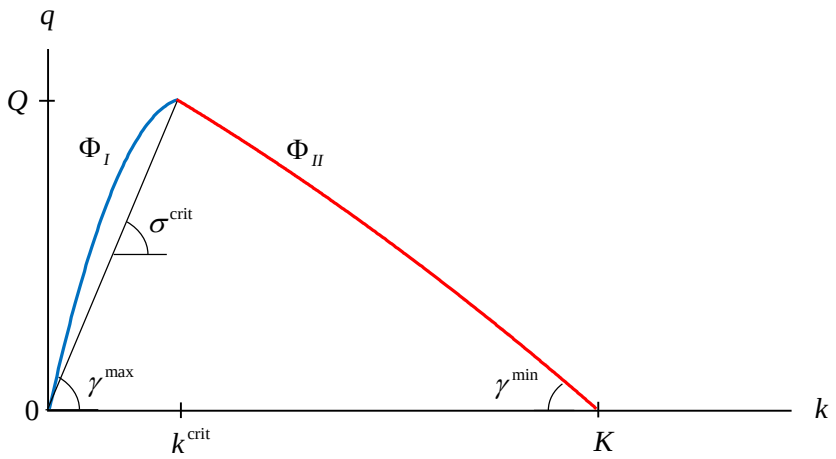


Figure 2: Dual-quadratic fundamental diagram

The dual-quadratic FD has five degrees of freedom, which makes it sufficiently flexible for most applications in transportation planning. The FDs of Greenshields (1935) and Smulders FD (1988), as well as the triangular FD are all special cases of the dual-quadratic FD as shown in Table 1. For example, Greenshields' diagram results if we set the jam density to $K=4Q/\gamma^{\max}$, the critical speed to $\sigma^{\text{crit}}=\frac{1}{2}\gamma^{\max}$, and the minimum wave speed to $\gamma^{\min}=-\gamma^{\max}$. This symmetric single-regime flux function has only two degrees of freedom, making it one of the least flexible flux functions used in practice.

Smulders' FD is a two-regime FD that combines the quadratic hypocritical branch of Greenshields' FD with the linear hypercritical branch. While this seems an attractive approach, it does not allow for specifying an exogenous critical speed. Our quadratic-linear FD is an extension of the Smulders' FD in which the user can input the critical speed as an additional degree of freedom. Since a linear hypercritical branch is often deemed sufficiently realistic, we argue that a quadratic-linear FD provides a good balance between realism and computational efficiency.

3. Model framework and definitions

Define graph $\mathcal{G}=(\mathcal{N},\mathcal{A})$ with the set of nodes $\mathcal{N}$ and the set of links $\mathcal{A}$. For each node $n\in\mathcal{N}$ let $\mathcal{A}_n^-$ and $\mathcal{A}_n^+$ denote the set of incoming and outgoing links, respectively. Each link $a\in\mathcal{A}$ is considered a homogenous road segment characterised by a link length L (km) and five calibration parameters of the FD as discussed in Section 2. Let $[0,T]$ (h) denote the simulation time period. Input into the DNL are path flow rates (veh/h) for every path between each origin-destination pair. We assume that these path flows are temporarily stationary. Note that in practice this is typically the case, since travel demand is often defined in terms of stationary flow rates per departure time interval and route choice is usually considered fixed per departure time interval. Further, we only consider a single vehicle class, such that the first-in-first-out property holds on the link level.

Table 1: Parameter values to generate different shapes of fundamental diagram

Fundamental diagram	Maximum wave speed	Capacity	Jam density	Critical speed	Minimum wave speed
	$\gamma^{\max}$	Q	K	σ^{crit}	$\gamma^{\min}$
Dual-quadratic	■	■	■	■	■
Quadratic-linear	■	■	■	■	$\frac{Q\sigma^{\text{crit}}}{2Q-K\sigma^{\text{crit}}}$
Smulders	■	■	■	$\frac{1}{2}\left(\gamma^{\max}+\sqrt{\gamma^{\max}\left(\gamma^{\max}-\frac{4Q}{K}\right)}\right)$	$\frac{Q\sigma^{\text{crit}}}{2Q-K\sigma^{\text{crit}}}$
Triangular	■	■	■	$\gamma^{\max}$	$\frac{Q\sigma^{\text{crit}}}{2Q-K\sigma^{\text{crit}}}$
Greenshields	■	■	$\frac{4Q}{\gamma^{\max}}$	$\frac{\gamma^{\max}}{2}$	$-\gamma^{\max}$

■ = calibration (input) parameter

LTM-based DNL models load these path flows onto the network by applying links models and node models. The general framework is shown in Figure 3. Path flow rates initially result in link inflow rates on the first link of each path, and are then propagated onto consecutive links via a node model. We first provide definitions of variables in the link model, and then we discuss the node model.

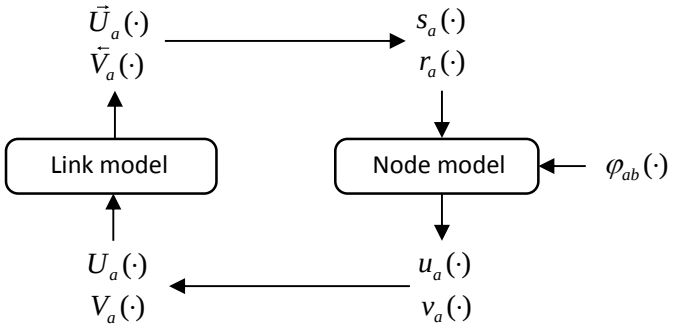


Figure 3: LTM-based DNL model framework

3.1 Link model definitions

Let $u_a(t)$ (veh/h) and $v_a(t)$ (veh/h) denote the inflow and outflow rates for link a and time instant t, respectively. We define the cumulative inflows $U_a(t)$ (veh) and cumulative outflows $V(t)$ (veh) for link a and time instant t as

$$U_a(t)=\int_0^t u_a(\omega)d\omega, \quad \text{and} \quad V_a(t)=\int_0^t v_a(\omega)d\omega. \tag{7}$$

Hence, the cumulative inflow (outflow) is the total number of vehicles that has entered (exited) a link up till time instant t. Due to the assumption of temporarily stationary path flow rates as input, the cumulative inflow on the first link of each path will therefore be piecewise linear. This means that $U_a(t)$ may not be differentiable at certain points in time. Therefore, in cases where $U_a(t)$ is differentiable it holds that $u_a(t)=dU_a(t)/dt$, while in other cases $u_a(t)\in\partial U_a(t)$, where $\partial U_a(t)$ denotes the sub- or

super-derivative.[1] Similarly, $v_a(t) = dV_a(t)/dt$ if $V(t)$ is differentiable in time instant *t*, and otherwise $v_a(t) \in \partial V_a(t)$.

Let $\vec{u}_a(t)$ (veh/h) denote the potential link *a* outflow rate at time instant *t* in case there are no outflow constraints, and let $\overleftarrow{v}_a(t)$ (veh/h) denote the potential inflow rate into link *a* at time instant *t* in case there are no inflow constraints. We use this notation to show that inflows and outflows influences each other, namely $\vec{u}_a(t)$ can be seen as a downstream projection of $u_a(t)$, and $\overleftarrow{v}_a(t)$ is an upstream projection of $v_a(t)$. Similar to the definitions in Equation (6) we can define the cumulative potential inflows $\vec{U}_a(t)$ (veh) the cumulative potential outflows $\overleftarrow{V}_a(t)$ (veh) for each link *a* and time instant *t* as

$$\vec{U}_a(t) = \int_0^t \vec{u}_a(\omega)d\omega, \quad \text{and} \quad \overleftarrow{V}_a(t) = \int_0^t \overleftarrow{v}_a(\omega)d\omega. \tag{8}$$

We illustrate these variables in Figure 4 where we consider a single link in which at time instant t_1 the outflow is restricted, leading to congestion and eventually spillback of the queue at the upstream link boundary at time instant t_2.

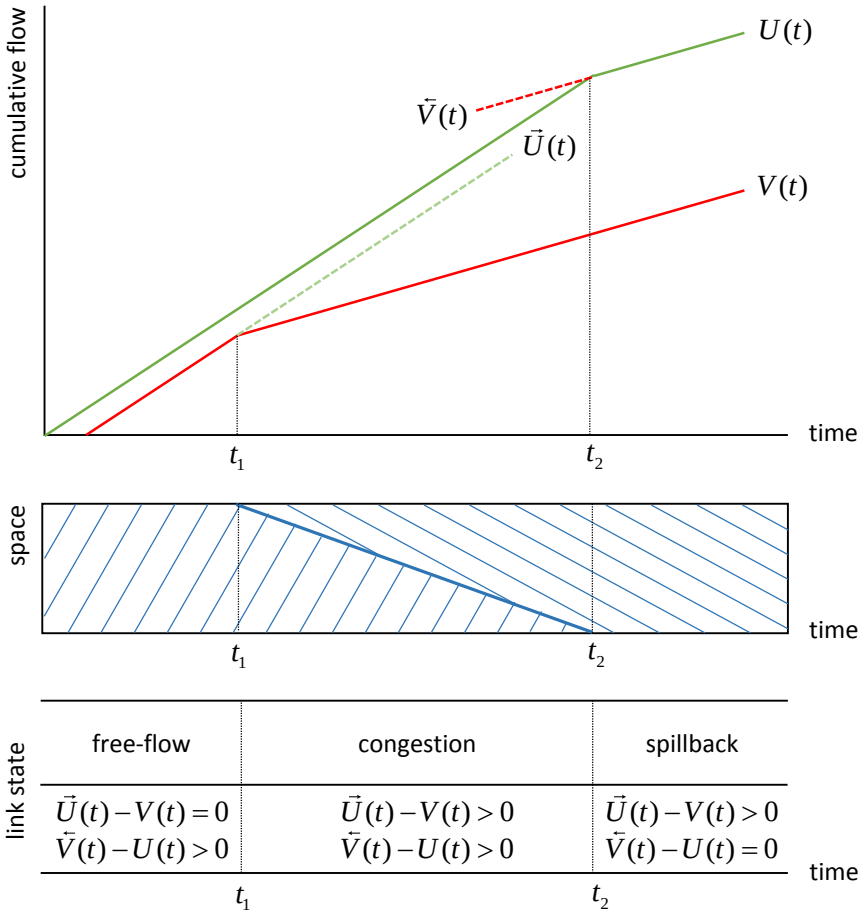


Figure 4: Cumulative potential and actual flows in different link states

[1] For a (locally) concave function *g*, super-derivative ∂g denotes the interval $\partial g = [\partial_+ g, \partial_- g]$, and for a (locally) convex function *g*, sub-derivative ∂g denotes the interval $\partial g = [\partial_- g, \partial_+ g]$, where $\partial_- g$ is the left-derivative and $\partial_+ g$ is the right-derivative. Clearly, if *g* is differentiable, then $\partial_- g = \partial_+ g$, and ∂g is a single value. While we use the same notation for sub- and super-derivative throughout the paper, the meaning remains the same, namely 'all values between the left- and right-derivative', and hence we believe this should not cause any confusion.

The top figure shows the cumulative (potential) flows, the middle figure illustrates the characteristic lines and shockwave in a space-time diagram, and the bottom part of the figure defines the different link states. It is clear in this figure that $\vec{U}_a(t)$ describes a downstream shift of the cumulative inflow, and $\overleftarrow{V}_a(t)$ can be seen as an upstream shift of the cumulative outflow. Since $\vec{U}_a(t)$ is the potential cumulative outflow and $V_a(t)$ is the actual cumulative outflow, the difference $\vec{U}_a(t) - V_a(t)$ can be defined as the excess demand and denotes the number of vehicles waiting on the link in a queue. Similarly, $\overleftarrow{V}_a(t) - U_a(t)$ is defined as the excess supply and represents the unused capacity.

We define three disjunct link states:

(i) *free-flow* – In this state, both the upstream and downstream link boundaries are in a hypocritical state, such that there is no congestion and no spillback, which means $\vec{U}_a(t) - V_a(t) = 0$ and by definition $\overleftarrow{V}_a(t) - U_a(t) > 0$;

(ii) *spillback* – In this state, both the upstream and downstream link boundaries are in a hypercritical state, such that there is congestion that is blocking back to upstream links, which means $\overleftarrow{V}_a(t) - U_a(t) = 0$ and by definition $\vec{U}_a(t) - V_a(t) > 0$;

(iii) *congestion* – In this state, the upstream link boundary is in a hypocritical state while the downstream link boundary is in a hypercritical state, which means that $\vec{U}_a(t) - V_a(t) > 0$ and $\overleftarrow{V}_a(t) - U_a(t) > 0$.

By using the term 'free-flow' we do not mean that vehicles drive at maximum speed, but we rather refer to any hypocritical traffic state (in which there may be some slowing down of traffic, but no queues). The LTM link model calculates $\vec{U}_a(t)$ and $\overleftarrow{V}_a(t)$, as well as $\vec{u}_a(t)$ and $\overleftarrow{v}_a(t)$. Equations for smooth nonlinear FDs are derived in Section 5.

3.2 Sending and receiving flow definitions

Based on the potential and actual cumulative inflows and outflows, we can calculate the sending flow rates $s_a(t)$ (veh/h) and receiving flow rates $r_a(t)$ (veh/h). The sending flow rate represents the demand for vehicles wanting to exit the link, while the receiving flow rate represents the supply for vehicles wanting to enter the link, and are given as

$$s(t) = \begin{cases} \vec{u}_a(t), & \text{if } \vec{U}_a(t) = V_a(t), \\ Q, & \text{otherwise,} \end{cases} \tag{9}$$

$$r(t) = \begin{cases} \overleftarrow{v}_a(t), & \text{if } \overleftarrow{V}_a(t) = U_a(t), \\ Q, & \text{otherwise.} \end{cases} \tag{10}$$

where $\vec{u}_a(t) \in \partial \vec{U}_a(t)$ and $\overleftarrow{v}_a(t) \in \partial \overleftarrow{V}_a(t)$. If the link is not in free-flow, then the sending flow rate is set to the link capacity. Similarly, if there is no spillback, then the receiving flow rate is set to the link capacity.

Discrete time models, such as proposed by Yperman et al. (2005) and Gentile (2010), consider a time step Δt. The average sending and receiving flow rates during time interval $[t, t+\Delta t)$ are then given as

$$s(t, t+\Delta t) = \min\left\{ \frac{1}{\Delta t}\left(\vec{U}(t+\Delta t) - V(t) \right),\ Q \right\}, \tag{11}$$

$$r(t, t+\Delta t) = \min\left\{ \frac{1}{\Delta t}\left(\overleftarrow{V}(t+\Delta t) - U(t) \right),\ Q \right\}. \tag{12}$$

Equations (11) and (12) converge to Equations (9) and (10) if $\Delta t \to 0$.

3.3 Node model definitions

Sending and receiving flow rates form a crucial connection between the link model and the node model. The node model considers for each node $n \in \mathcal{N}$ the demand and supply for all incoming links $a \in \mathcal{A}_n^-$ and all outgoing links $b \in \mathcal{A}_n^+$. In order to know which vehicle flows interact and compete with each other for capacity, sending flow rates $s_a(t)$ need to be decomposed into direction-specific sending flow rates representing the demand for a movement from link *a* to link *b*. Let $\varphi_{ab}(t)$ denote the proportion of the vehicle flow from link *a* that would like to enter link *b* at time instant *t*. It holds that $\sum_b \varphi_{ab}(t) = 1$. Then the demand for a movement from link *a* to link *b* is $\varphi_{ab}(t) s_a(t)$.

Calculation of $\varphi_{ab}(t)$ depends on the type of DNL that is considered. In case of a path-based DNL, $\varphi_{ab}(t)$ is calculated during the simulation by keeping track of path-specific link inflow and outflow rates. A path-based DNL is a computationally demanding multi-commodity model, which for large scale networks may be infeasible (the number of commodities is equal to the total number of paths in the network). Instead, one may consider a simplified multi-commodity model in which one only keeps track of destination-specific inflows and outflows, which will significantly reduce the number of commodities considered at the cost of some loss in path consistency. A more computationally efficient type of DNL calculates fixed turn proportions for each departure time period at the beginning of the simulation based on the path flow rates. This leads to a specific multiclass model in which we need to distinguish between vehicles that departed during different time periods. Such a multiclass model is significantly more efficient than the earlier mentioned multi-commodity models, however comes at a greater loss of path consistency. Since our focus in this paper is on link models, we do not specify which type of DNL model we consider, but we note that all types can be considered in our framework (under the assumption of a single vehicle type).

Given the demand and supply at a node, movement specific outflow rates $v_{ab}(t)$ are then determined by a node model, which is given by the following general form:

$$[v_{ab}(t)]_{a \in \mathcal{A}_n^-, b \in \mathcal{A}_n^+} = \Gamma_n \left(s_a(t), r_a(t), \varphi_{ab}(t), \quad a \in \mathcal{A}_n^-, b \in \mathcal{A}_n^+ \right). \tag{13}$$

LTM models work with any function $\Gamma_n(\cdot)$ that satisfies the requirements outlined in Tampère et al. (2011). Smits et al. (2015) identified a family of models that satisfy these requirements. In case of a through node with a single incoming link *a* and a single outgoing link *b*, this function is simply $\Gamma_n \left(s_a(t), r_b(t) \right) = \min\{ s_a(t), r_b(t), 1 \}$.

Link inflow rates can be determined as $u_b(t) = \sum_{a \in \mathcal{A}_n^-} v_{ab}(t)$ for each link $b \in \mathcal{A}_n^+$, and link outflow rates can be computed as $v_a(t) = \sum_{b \in \mathcal{A}_n^+} v_{ab}(t)$ for each link $a \in \mathcal{A}_n^-$.

In the remainder of the paper we focus on a single link and therefore omit subscript *a* for notational convenience.

4. Link model with a nonlinear differentiable concave fundamental diagram

LTM formulations can be obtained by solving the Hamilton-Jacobi problem formulation of the kinematic wave model using the Lax-Hopf formula. We refer to Appendix A for a quick introduction into the relevant theory.

Consider a general two-regime concave FD as in Equation (1), where we assume that $\Phi_I(k)$ is strictly increasing, and $\Phi_{II}(k)$ is strictly decreasing. Further, we assume that $\Phi_I(k)$ and $\Phi_{II}(k)$ are differentiable over their entire domain. Let γ_I^{crit} and $\gamma_{II}^{\text{crit}}$ denote the wave speeds at capacity in the hypocritical and hypercritical branch, respectively. We define $\gamma_I : [0, k^{\text{crit}}]$ (veh/km) $\rightarrow [\gamma_I^{\text{crit}}, \gamma^{\max}]$ (km/h) and $\gamma_{II} : [k^{\text{crit}}, K]$ (veh/km) $\rightarrow [\gamma^{\min}, \gamma_{II}^{\text{crit}}]$ (km/h) as the wave speeds corresponding to a hypocritical or hypercritical traffic state with density *k*, respectively, which are given by

$$\gamma_I(k) = \frac{d\Phi_I(k)}{dk}, \quad \text{and} \quad \gamma_{II}(k) = \frac{d\Phi_{II}(k)}{dk}. \tag{14}$$

Further, since we assumed that the hypocritical (hypercritical) branch is strictly increasing (decreasing), the flux functions can be inverted. Hence, instead of representing traffic states by a density we can represent traffic states by flow rates, which is more common in LTM (since boundary conditions are provided in terms of flow rates). Therefore, we can define $\sigma_I(q) = q / \Phi_I^{-1}(q)$, $\sigma_{II}(q) = q / \Phi_{II}^{-1}(q)$, $\gamma_I(q) = \gamma_I\left(\Phi_I^{-1}(q)\right) = (d\Phi_I^{-1}(q)/dq)^{-1}$, and $\gamma_{II}(q) = \gamma_{II}\left(\Phi_{II}^{-1}(q)\right) = (d\Phi_{II}^{-1}(q)/dq)^{-1}$.

Since we consider a two-regime FD, the Lax-Hopf formula in Equation (A.4) can be applied separately for traffic states propagating forwards and backwards. Hypocritical traffic states travel from the upstream to the downstream link boundary, while hypercritical traffic states travel from the downstream to the upstream link boundary. We note that in Equation (A.4) $N(x',t)$ equals $U(t)$ if $x' = 0$ and equals $V(t)$ if $x' = L$. Further, $N(x,t)$ equals $\overleftarrow{V}(t)$ if $x = 0$ and equals $\vec{U}(t)$ if $x = L$, referring to potential cumulative inflows and outflows instead of actual cumulative inflows and outflows since there may be possible outflow and inflow constraints.

4.1 Calculating potential cumulative outflows

We first look at the situation where the upstream link boundary is in a hypocritical state. Then the Lax-Hopf formula can be written as[2]

$$\vec{U}(t) = \min_{t'}\left\{U(t') + (t-t')\Phi_I^*\left(\frac{L}{t-t'}\right)\right\}, \quad \Phi_I^*(w) = \max_{k\in[0,k^{\text{crit}}]}\left\{\left(\sigma_I(k) - w\right)k\right\}, \tag{15}$$

We can rewrite Equation (15) so that it minimises over (wave) speed w instead of minimising over t'. It holds that $w = L/(t-t')$, such that substituting $t' = t - L/w$ in Equation (15) results in

$$\vec{U}(t) = \min_{w\in[\gamma_I^{\text{crit}},\gamma^{\max}]}\left\{U\left(t - \frac{L}{w}\right) + \frac{L}{w}\Phi_I^*(w)\right\}. \tag{16}$$

Due to our assumption that Φ_I is strictly increasing, it holds that $\gamma_I^{\text{crit}} > 0$. Therefore, dividing by w over the strict positive range $[\gamma_I^{\text{crit}}, \gamma^{\max}]$ is not problematic.

Define $\xi_I(w) = (L/w)\Phi_I^*(w)$ (veh), which can be rewritten as

$$\xi_I(w) = \max_{k\in[k^{\text{crit}},K]}\left\{Lk\left(\frac{\sigma_I(k)}{w} - 1\right)\right\}, \tag{17}$$

where $\xi_I(w)$ (veh) represents the number of vehicles that an observer passes while traversing the link downstream at kinematic wave speed w.

This density based formulation can also be replaced by a flow based formulation. Since Φ_I is strictly monotone, it is invertible, such that we can also write the following equivalent flow based formulation:[3]

[2] Where we have replaced the infimum with a minimum and the supremum with a maximum since all functions are continuously differentiable and bounded.

[3] LTM typically expresses all variables in terms of flow (see e.g. Yperman et al., 2005; Gentile, 2010), but depending on the functional form of the FD one could choose for a density based formulation if that simplifies calculations. Whether a density or a flow based formulation is preferred depends on the functional formulation of the flux functions. In case of our dual-quadratic FD, functions $\Phi_I(k)$ and $\Phi_{II}(k)$ are mathematically easier to deal with than $\Phi_I^{-1}(q)$ and $\Phi_{II}^{-1}(q)$.

$$\xi_I(w) = \max_{q\in[0,Q]} \left\{ Lq\left(\frac{1}{w} - \frac{1}{\sigma_I(q)} \right) \right\}. \tag{18}$$

Now we can write the general formula that propagates hypocritical states from the upstream to the downstream link boundary as

$$\vec{U}(t) = \min_{w\in[\gamma_I^{\text{crit}},\gamma^{\max}]} \left\{ U\left(t - \frac{L}{w} \right) + \xi_I(w) \right\}. \tag{19}$$

It holds that $\xi_I(w)$ is non-negative and a decreasing function, with $\xi_I(\gamma^{\max}) = 0$. Since $U(t)$ is a monotonically increasing function, it holds that $U(t - L/w)$ is increasing in w. Therefore, $U(t - L/w)$ and $\xi_I(w)$ work in opposite directions when minimizing with respect to wave speed w.

Since Φ_I is assumed to be differentiable, the minimum in Equation (19) is achieved at the (unique) density $k^* \in [0, k^{\text{crit}}]$ for which holds that $w = \gamma_I(k^*)$. Therefore, instead of minimising over wave speeds in Equation (19), we can minimise over densities and simply set $w = \gamma_I(k)$. This results in the following density based formulation for cumulative potential outflows,

$$\vec{U}(t) = \min_{k\in[0,k^{\text{crit}}]} \left\{ U\left(t - \frac{L}{\gamma_I(k)} \right) + \xi_I(k) \right\}, \text{ with } \xi_I(k) = Lk\left(\frac{\sigma_I(k)}{\gamma_I(k)} - 1 \right). \tag{20}$$

An equivalent flow based formulation is

$$\vec{U}(t) = \min_{q\in[0,Q]} \left\{ U\left(t - \frac{L}{\gamma_I(q)} \right) + \xi_I(q) \right\}, \text{ with } \xi_I(q) = Lq\left(\frac{1}{\gamma_I(q)} - \frac{1}{\sigma_I(q)} \right). \tag{21}$$

Equation (21) states that the cumulative potential outflow is determined by the cumulative inflow exactly $L/\gamma_I(q)$ earlier (i.e., following the characteristic line of traffic state q) plus a correction for the number of vehicles 'overtaking' the characteristic wave during this period (since vehicles drive at a higher speed than wave speed $\gamma_I(q)$ in the same direction).

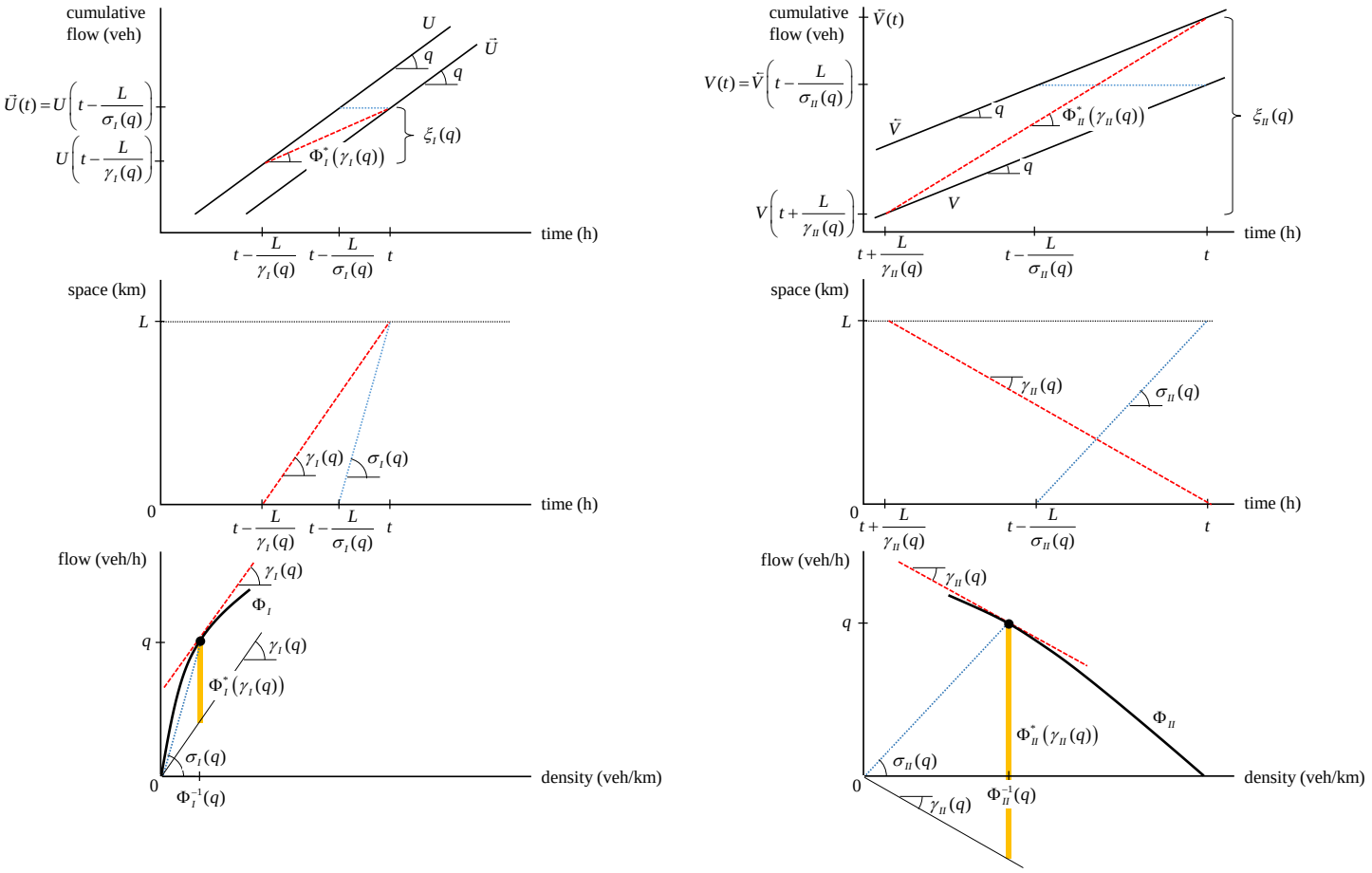
The corresponding potential outflow rate at time instant t, i.e., $\vec{u}(t) \in \partial\vec{U}(t)$, can alternatively (and more easily) be obtained by finding the argument of the minimisation in Equation (21), in other words,

$$\vec{u}(t) = \arg\min_{q\in[0,Q]} \left\{ U\left(t - \frac{L}{\gamma_I(q)} \right) + \xi_I(q) \right\}. \tag{22}$$

Note that this flow rate need not be unique.

Figure 5(a) graphically shows the various relationships for a simple case in which the upstream link boundary is in a stationary hypocritical state with flow rate q. The top graph shows the cumulative link inflow $U(t)$ and potential cumulative outflow $\vec{U}(t)$, the middle graph shows a space-time diagram, and the bottom graph shows the traffic state (i.e., flow q) on the hypocritical branch of the FD. Blue (dotted) lines relate to propagation of vehicles, whereas red (dashed) lines relate to propagation of traffic states. In the bottom diagram, at traffic flow q the slope of the blue line indicates the speed at which vehicles propagate, i.e. $\sigma_I(q)$, whereas the slope of the red line indicates the (wave) speed w at which traffic states propagate, which is the tangent of the FD at flow q. In the middle diagram, a vehicle that can potentially exit the link at time instant t entered the link at time instant $t - L/\sigma_I(q)$. In other words, the link travel time is at least $L/\sigma_I(q)$ depending on whether there are outflow constraints. The traffic state (i.e., flow rate q) travels at speed $\gamma_I(q)$ and therefore takes a time of $L/\gamma_I(q)$ to arrive at the end of the link at time instant t (noting that traffic states never travel faster than vehicles). In the top diagram, since FIFO holds, it is shown that $U\left(t - L/\sigma_I(q)\right) = \vec{U}(t)$. Furthermore, the difference between $\vec{U}(t)$ and $U(t - L/\gamma_I(q))$ is equal to $\xi_I(q)$. A nice interpretation of $\Phi_I^*(\gamma_I(q))$, the Legendre-Fenchel

transform of the hypocritical branch of the FD at traffic state q, is that it represents the slope of the line that connects this traffic state between the cumulative link inflows and outflows.



(a) Hypocritical states (b) Hypercritical states

Figure 5: Relationships in (a) hypocritical and (b) hypercritical states in link transmission models

4.2 Calculating potential cumulative inflows

A similar derivation for the situation where the downstream link boundary is in a hypercritical state leads to the following equation for the potential cumulative inflow (see Appendix B):

$$\overleftarrow{V}(t) = \min_{q\in[0,Q]}\left\{V\left(t+\frac{L}{\gamma_{II}(q)}\right)+\xi_{II}(q)\right\}, \quad \text{with} \quad \xi_{II}(q) = Lq\left(\frac{1}{\sigma_{II}(q)}-\frac{1}{\gamma_{II}(q)}\right), \tag{23}$$

where $\xi_{II}(q)$ (veh) represents the number of vehicles that an observer passes while traversing the link upstream at kinematic wave speed $\gamma_{II}(q)$. Assuming that both upstream and downstream traffic conditions are hypercritical, Equation (23) states that the cumulative potential inflow is determined by the cumulative outflow exactly $L/\gamma_{II}(q)$ earlier (i.e., following the characteristic line of traffic state q) plus a correction for the number of vehicles that 'overtake' the characteristic wave speed $\gamma_{II}(q)$ during this period because they are driving at a higher speed (in the opposite direction).

The corresponding potential inflow rate at time instant t, i.e., $\overleftarrow{v}(t) \in \partial\overleftarrow{V}(t)$, is alternatively (and more easily) obtained via the argument of the minimisation in Equation (23),

$$\bar{v}(t) = \arg\min_{q\in[0,Q]} \left\{ V\left(t + \frac{L}{\gamma_{II}(q)} \right) + \xi_{II}(q) \right\}. \tag{24}$$

Again, this argument need not be unique.

Figure 5(b) illustrates the relationships between the variables in case the downstream link boundary is in a stationary hypercritical state with flow rate q. In the bottom diagram, at traffic flow q the slope of the blue (dotted) line indicates the speed at which vehicles propagate, i.e. $\sigma_{II}(q)$, whereas the slope of the red (dashed) line indicates the (wave) speed w at which traffic states propagate (backwards), which is the tangent of the FD at flow q. In the middle diagram, a vehicle that arrived at the end of the link at time instant t would have entered the link at the latest at time instant $t - L/\sigma_{II}(q)$. The traffic state (i.e., flow rate q) travels with (negative) speed $\gamma_{II}(q)$ and therefore takes a time of $-L/\gamma_{II}(q)$ to arrive at the beginning of the link at time instant t. Because of FIFO, it holds that $\bar{V}\left(t - L/\sigma_{II}(q)\right) = V(t)$ in the top diagram, and the difference between $\bar{V}(t)$ and $V(t + L/\gamma_{II}(q))$ is equal to $\xi_{II}(q)$. The Legendre-Fenchel transform of the hypercritical branch of the FD at traffic state q, denoted by $\Phi^*_{II}(\gamma_{II}(q))$, represents the slope of the line that connects this traffic state between the cumulative link outflows and inflows.

4.3 Relationships with existing link transmission models

Equations (21) and (23) are consistent with the expressions proposed in Gentile (2010), and lead to physically meaningful (entropic) solutions to the kinematic wave model. The minimisation over all possible traffic states is necessary in case of a nonlinear FD, which also takes expansion fans into account.

In case of a linear hypocritical branch (e.g., in the triangular FD), then $\xi_I(q) = 0$ and $\gamma_I(q) = \gamma^{\max}$, such that Equation (21) simplifies to

$$\vec{U}(t) = U\left(t - \frac{L}{\gamma^{\max}} \right). \tag{25}$$

When the hypercritical branch is linear (e.g., in the triangular, Smulders', and quadratic-linear FD), then $\xi_{II}(q) = LK$ and $\gamma_{II}(q) = \gamma^{\min}$, such that Equation (23) simplifies to

$$\bar{V}(t) = V\left(t + \frac{L}{\gamma^{\min}} \right) + LK. \tag{26}$$

These equations are equivalent to the expressions proposed in Yperman et al. (2005), Jin (2015), and Han (2016b). Clearly, the main advantage of using linear branches is that the minimisation over all traffic states can be avoided, which also removes expansion fans.

4.4 Numerical example

In this section we describe an example how to calculate $\vec{U}(t)$. We will use this example throughout the paper to illustrate certain equations and concepts. Calculation of $\bar{V}(t)$ works in a similar fashion and is therefore not further discussed.

Example 1 – Consider a link with $L = 1$ (km), $Q = 2000$ (veh/h). $\gamma^{\max} = 120$ (km/h), $\sigma^{\text{crit}} = 80$ (km/h), and $K = 180$ (veh/km), and assume an FD with a quadratic hypocritical branch, see Section 2. Assume that the link inflow rate is stationary at 200 veh/h for a long time, but that at $t = 50$ s the inflow rate increases to 1800 veh/h, and drops to 1000 veh/h at $t = 100$ s. Using Equations (21) and (22) to construct the potential cumulative outflow and the potential outflow rate results in Figure 6. Figure 6(a) shows the cumulative link inflows (solid blue line) and cumulative potential link outflows (dashed red line)

over time, and Figure 6(b) shows the link inflow rates (solid blue line) and potential outflow rates (dashed red line) over time. The dotted lines that connect the large dots on the cumulative inflow and cumulative potential outflow curves show examples of how flow rates propagate forward, and reflect characteristic lines in which traffic states remain constant. The increase in link inflow rate leads to fanning and therefore to potential outflow rates that gradually increase from 200 veh/h to 1800 veh/h. Fanning does not occur when the flow decreases, i.e. the decrease in the link inflow rate from 1800 veh/h to 1000 veh/h leads to the same drop in the potential outflow rate. This example also shows that the potential outflow rate need not be unique. The potential outflow rate at $t = 150$ s is essentially equal to two different inflow rates, namely 1800 veh/h and 1000 veh/h, as shown by the large black dots in Figure 6(a). Figure 6(c) shows a space-time diagram in which the characteristic lines have been plotted. The expansion fan is indicated by the grey area in which flows rates continuously increase from 200 to 1800 veh/h, and the solid black line indicates a shockwave that separates flow rates 1800 veh/h and 1000 veh/h. ■

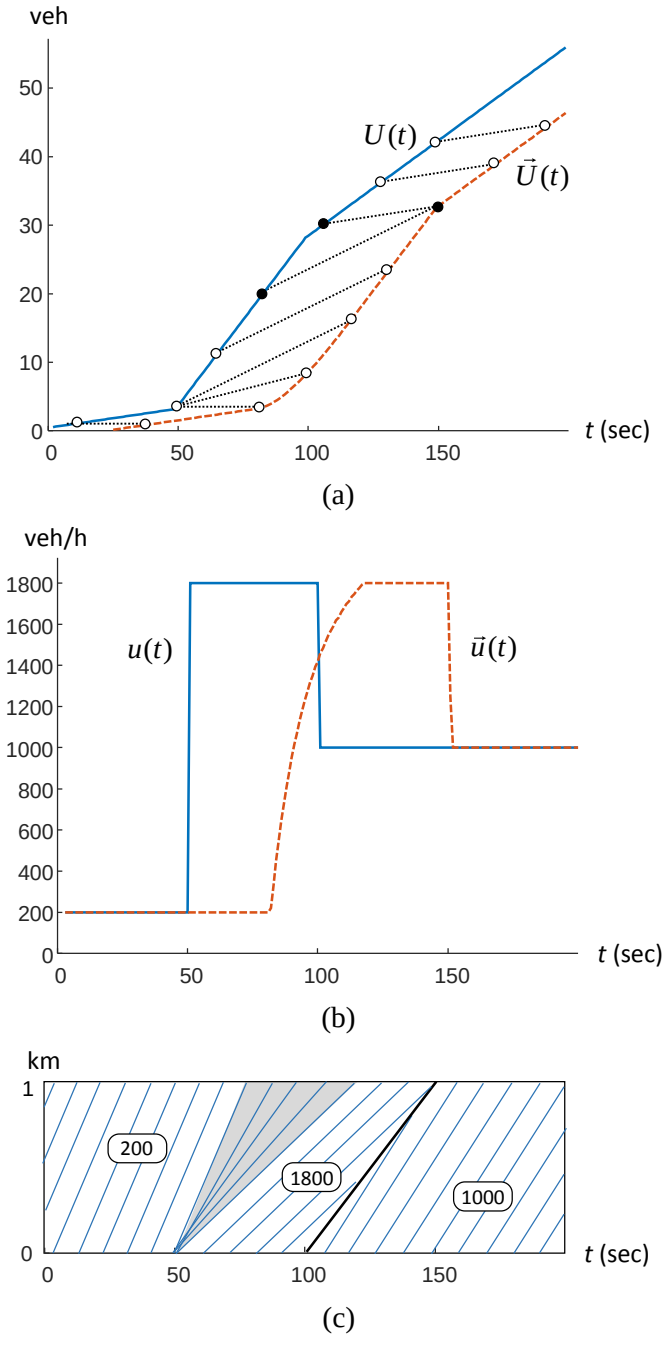


Figure 6: Numerical example 1

Due to the need to minimise over all traffic states as shown in Equations (21) and (22), these equations are not easy to solve over a given time period when using nonlinear FDs, mainly due to the presence of expansion fans. Therefore, we solved Example 1 numerically using time discretisation using $\Delta t = 1$ s, and as such an approximate entropic solution was obtained (which is in line with Gentile, 2010).

5. On-the-fly linearization of the fundamental diagram to simplify expansion fans

As we have seen, even if the cumulative inflows in Example 1 are piecewise linear, the resulting potential cumulative outflow is in general nonlinear due to the presence of expansion fans when the flow rate increases. However, if we would have adopted a piecewise linear FD, fanning would have simplified such that the potential outflow rate would not continuously increase but rather increase in discrete steps, resulting in a piecewise potential cumulative flows as output. This allows finding an exact solution under a fixed simplified FD.

We provided several arguments in Section 2 why we prefer to use nonlinear differentiable branches over fixed piecewise linear branches in the FD, especially for hypocritical traffic states. Therefore, in this section we propose a method in which we use a nonlinear FD as input, but propose on-the-fly linearization to simplify fanning. The distinction between using a fixed piecewise linear FD and using on-the-fly linearization may sound subtle, but is important. In the former case the user provides simplified input (in the form of a linear or piecewise linear FD) into the kinematic wave model in order to find an entropic solution, while in the latter case the user provides non-simplified input (in the form of a smooth nonlinear FD) into a simplified kinematic wave model in order to find a non-entropic solution. The latter allows more flexibility by dynamically linearizing the FD as appropriate during the simulation and is able to maintain the advantages of a nonlinear FD (i.e., realism in most cases, uniqueness of traffic equilibrium, ease of calibration). If we would have used a piecewise linear FD in Example 1, the entire potential cumulative outflow curve would have been modified. In contrast to the on-the-fly linearization that we propose, only the fanning part would be modified. This is achieved by maintaining the points on the nonlinear FD that correspond to flow rates that are input into the link model, while we approximate only the points in between. These points in between are only used to describe expansion fans and therefore our linearization methods do not affect other flow rates.

We propose two types of on-the-fly linearization, discussed in the next two subsections.

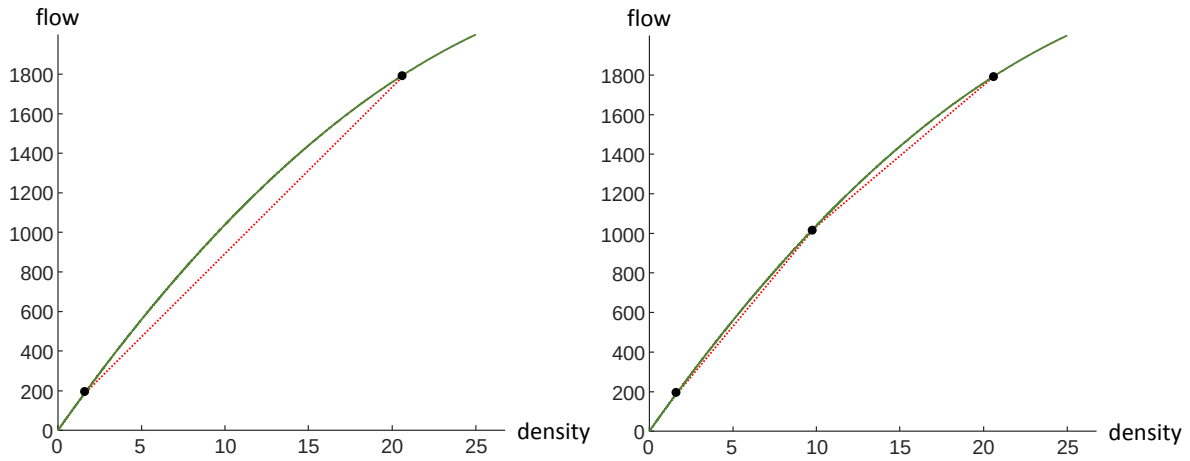
5.1 Inner linearization

The first method we call inner linearization, which is essentially basic interpolation that results in an approximation that lies below the FD curve. The simplest inner linearization for Example 1 is illustrated in Figure 7(a) in which we use linear interpolation between the two traffic states that are given by the input, 200 veh/h and 1800 veh/h, indicated by the red dotted line. We call this 1-step inner linearization in which we move in one step from 200 to 1800 veh/h.

Figure 7(b) shows the case in which we have used one additional point between the two existing flow rates (in this example we have put this point in the middle at 1000 veh/h, but we could have chosen this point differently). This linearization with two segments we call a 2-step inner linearization. In general, in a *D*-step inner linearization we interpolate between $D+1$ points on the FD, with $D \geq 1$.

Mathematically, let $\Phi_I^{-1,\text{in}}(q \mid q_0,\ldots,q_D)$ denote the *D*-step inner linearization of the inverse flux function of the hypocritical branch using flow rates $q_0,\ldots,q_D$ (in increasing order). As we will see, this results in expansion fans in which the flow rate increases in *D* steps. This function is defined as

$$\Phi_I^{-1,\text{in}}(q \mid q_0,\ldots,q_D) = \begin{cases} \dfrac{\Phi_I^{-1}(q_{d+1})(q-q_d) - \Phi_I^{-1}(q_d)(q-q_{d+1})}{q_{d+1}-q_d}, & \text{if } q_d \leq q \leq q_{d+1}, \text{for } d = 0,\ldots,D-1, \\ \Phi_I^{-1}(q), & \text{otherwise.} \end{cases} \tag{27}$$



(a) 1-step inner linearization (b) 2-step inner linearization

Figure 7: On-the-fly inner linearization of the FD to simplify fanning

The slope of each these interpolated lines (dotted in red in Figure 7) is the inverse of the hypocritical shockwave speed $\eta_I(q_d, q_{d+1})$ between two consecutive traffic states q_d and q_{d+1}, which in general for any two flow rates q_1 and q_2 is given as

$$\eta_I(q_1, q_2) = \frac{q_2 - q_1}{\Phi_I^{-1}(q_2) - \Phi_I^{-1}(q_1)}. \tag{28}$$

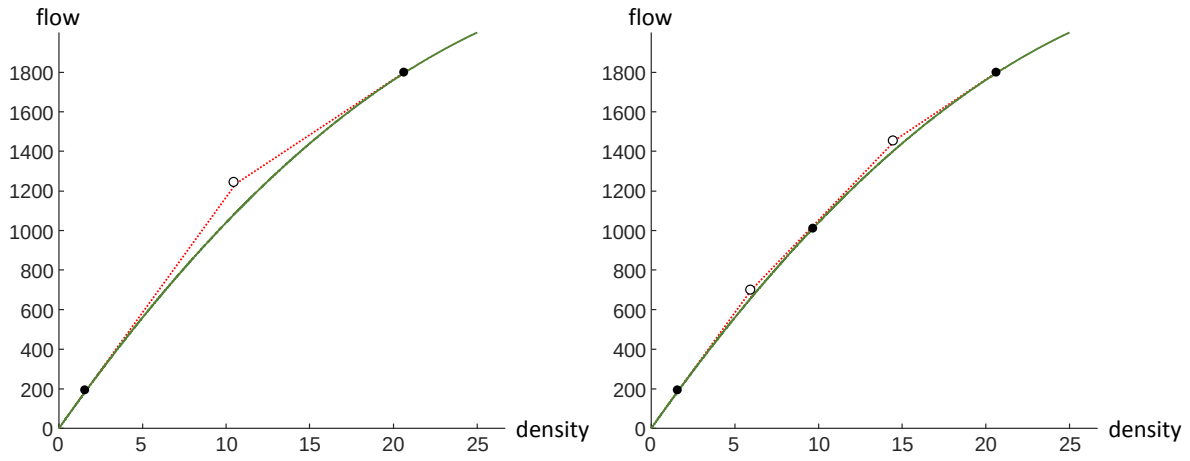
Similarly, for the hypercritical branch we can define similar equations by simply replacing subscript *I* with subscript *II*.

Since the inner linearization lies below the FD, this means that the vehicle speeds in the hypocritical branch will be underestimated, while they will be overestimated for traffic states in the hypercritical branch. This over- or underestimation can be reduced by increasing the number of steps *D*. The original nonlinear FD would be obtained in the limit when $D \to \infty$, therefore setting *D* is a matter of finding a balance between accuracy and computational efficiency.

5.2 Outer linearization

The second method we call outer linearization in which we use extrapolation adopting the wave speeds of traffic states on the FD, resulting in an approximation that lies above the FD. The simplest outer linearization for Example 1 is shown by the red dotted line in Figure 8(a) where we extrapolate the FD using the wave speeds in flow rates 200 veh/h and 1800 veh/h. We call this a 2-step outer linearization in which we move from 200 veh/h to 1800 veh/h via a new point indicated with a white dot. In contrast to the inner linearization, we now not only preserve relevant traffic states on the FD, but we also preserve their corresponding wave speeds. Figure 8(b) shows a 3-step outer linearization based on one additional point. In general we can introduce a *D*-step outer linearization using flow rates $q_1, \ldots, q_D$ on the FD (in increasing order), which results in expansion fans in which the flow rate increases with *D* steps. Note that $D \geq 2$, that is, a 1-step outer linearization does not exist.

Let $\Phi_I^{-1,\text{out}}(q \mid q_1, \ldots, q_D)$ denote the *D*-step outer linearization of the inverse flux function of the hypocritical branch, which is defined mathematically as



(a) 2-step outer linearization (b) 3-step outer linearization

Figure 8: On-the-fly outer linearization of the FD to simplify fanning

$$\Phi_I^{-1,\text{out}}(q \mid q_1,\ldots,q_D) = \begin{cases} \dfrac{q-q_1}{\gamma_I(q_1)} + \Phi_I^{-1}(q_1), & \text{if } q_1 \le q \le \phi_I(q_1,q_2), \\ \dfrac{q-q_d}{\gamma_I(q_d)} + \Phi_I^{-1}(q_d), & \text{if } \phi_I(q_{d-1},q_d) \le q \le \phi_I(q_d,q_{d+1}), \text{ for } d = 2,\ldots,D-1, \\ \dfrac{q-q_D}{\gamma_I(q_D)} + \Phi_I^{-1}(q_D), & \text{if } \phi_I(q_{D-1},q_D) \le q \le q_D, \\ \Phi_I^{-1}(q), & \text{otherwise,} \end{cases} \tag{29}$$

where the flow rate at which the two segments meet, denoted by $\phi_I(q_d,q_{d+1})$, is given by

$$\phi_I(q_d,q_{d+1}) = \frac{\gamma_I(q_d)\left(q_{d+1} - \gamma_I(q_{d+1})\Phi_I^{-1}(q_{d+1})\right) - \gamma_I(q_{d+1})\left(q_d - \gamma_I(q_d)\Phi_I^{-1}(q_d)\right)}{\gamma_I(q_d) - \gamma_I(q_{d+1})}. \tag{30}$$

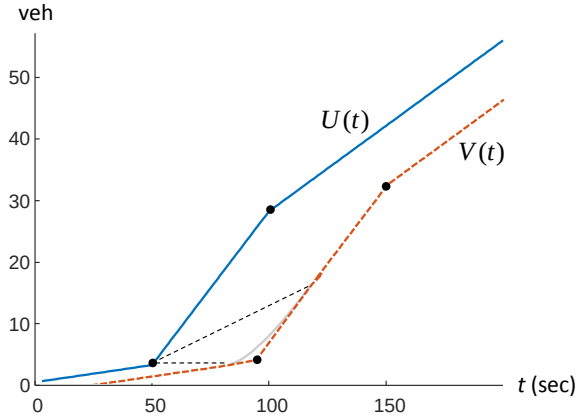
For Example 1, we can calculate that $\phi_I(200,1800) = 1290$ veh/h. For outer linearization of the hypercritical branch we replace subscript *I* with subscript *II* in Equations (28)–(30). Similar to the inner linearization, this outer linearization converges to the nonlinear FD if $D \to \infty$.

5.3 Simplified fanning

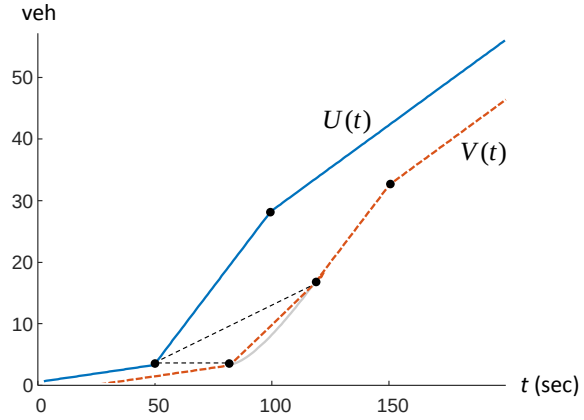
In order to understand what our linearization methods achieve in terms of simplifying expansion fans, we again look at Example 1.

First, consider the on-the-fly 1-step inner linearization of the FD as shown in Figure 8(a). Since the FD is no longer differentiable during the time period that fanning occurs, we cannot apply Equations (21) and (22). Instead, we need to go back to Equations (18) and (19) that holds for any concave FD. Considering the expansion fan due to the flow increase at $t' = 50$ s, for all wave speeds $\gamma_I(200) < w < \eta_I(200,1800)$ the Legendre-Fenchel transform $\Phi_I^*(w)$ achieves its maximum at a flow rate of 200 veh/h, while for all wave speeds $\eta_I(200,1800) < w < \gamma_I(1800)$ its maximum is achieved at a flow rate of 1800 veh/h. This fact is most easily observed using the graphical illustration in Figure A.1 in Appendix A. As a result, the outflow rate remains 200 veh/h for some time even after $t' + L/\gamma_I(200)$ and jumps to 1800 veh/h at time instant $t' + L/\eta_I(200,1800)$. This is shown graphically in Figure 9(a). The black dots in the figure indicate changes in flow rate, and it is clear that fanning using on-the-fly inner linearization is greatly simplified, and the cumulative outflow curve becomes piecewise linear. It is further clear that vehicle travel times are somewhat over-estimated in the expansion fan. As shown in Figure 9(b), there are only three outflow rates, namely 200, 1000, and 1800

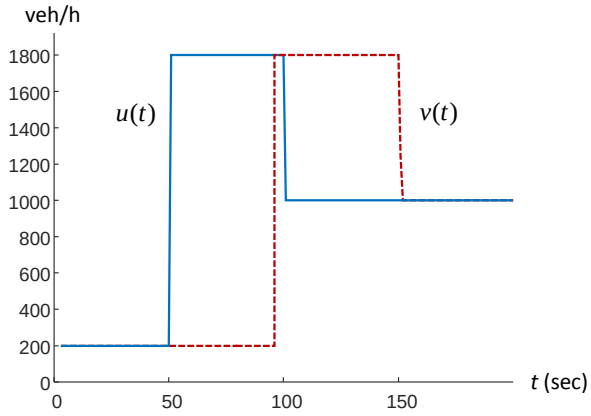
veh/h, which are identical to the inflow rates. Figure 9(c) shows that an *inner* wave (equal to the shockwave) is used to separate flow rates in the expansion fan. We note that 1-step inner linearization is consistent with what is sometimes referred to as shockwave theory or shockwave analysis (see May, 1990), which simplifies kinematic wave theory by only considering shockwaves when flow rates change. Our *D*-step inner linearization is therefore an extension to shockwave theory in which we use multiple shockwaves to replace the expansion fan. As we will show in Section 6, this no longer leads to an entropic solution but rather to a non-entropic solution of the kinematic wave model. If D becomes large, then the non-entropic solution approaches the entropic solution.



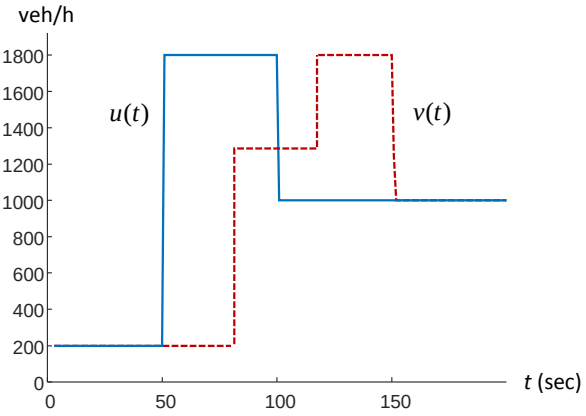
(a) Cumulatives with 1-step inner linearization



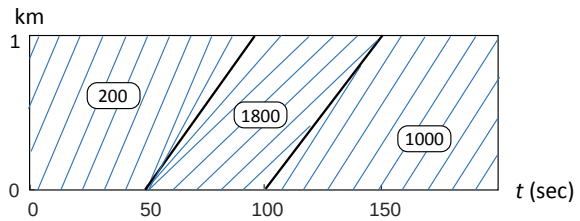
(d) Cumulatives with 2-step outer linearization



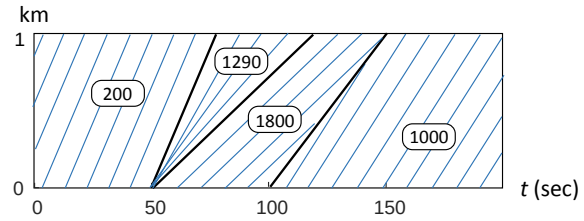
(b) Flow rates with 1-step inner linearization



(e) Flow rates with 2-step outer linearization



(c) Space-time diagram 1-step inner linearization



(f) Space-time diagram 2-step outer linearization

Figure 9: Simplified fanning in Example 1

Next, we consider the on-the-fly 2-step outer linearization of the FD as shown in Figure 9(d). Now $\Phi_I^*(w)$ achieves its maximum at $\phi_I(200,1800)$ as defined in Equation (30) for all wave speeds

$\gamma_I(200) < w < \gamma_I(1800)$. This means that the outflow rate becomes $\phi_I(200,1800)$ after $t' + L/\gamma_I(200)$ and increases to 1800 veh/h at $t' + L/\gamma_I(1800)$. This situation is shown in Figure 9(e). In contrast to the inner linearization, vehicle travel times are now under-estimated during fanning. Comparing Figure 9(e) with 9(b), an additional step in the outflow rate at 1290 veh/h can now be observed. This is further shown in Figure 9(f) that illustrates that the two *outer* waves (corresponding to traffic states 200 and 1800 veh/h) separate flow rates in the expansion fan. *D*-step outer linearization of the FD is consistent with the resolution scheme proposed by Henn (2003) in which the expansion fan is simplified by considering a finite set of traffic states (in his case densities) separated by multiple waves starting with the fastest wave and ending with the slowest wave. As Henn (2003) points out, this yields a non-entropic solution to the kinematic wave theory and approximates, the more difficult to find, entropic solution, which can be found more accurately by increasing the resolution (i.e., the number of on-the-fly linearization steps *D*).

We emphasize that *D* does not need to be fixed throughout the simulation, but can vary dynamically depending on the increase in flow rate. Note only the number of steps, but also the location of the intermediate points on the FD can be chosen differently. A simple rule could be that for an increase in flow rate from q_1 to q_2 we select the smallest number of additional points that are equidistant with respect to flow rates (or alternatively equidistance in terms of density or wave speeds) for which holds that $\frac{1}{D}(q_2 - q_1) \leq \Delta q$, where Δq is a user-defined upper bound on the maximum allowed 'jump' between flow rates. The 2-step inner linearization in Figure 7(b) would for example be the result of setting $\Delta q = 1000$ veh/h in Example 1. If the maximum allowed jump would be 300 veh/h then a 6-step inner linearization would be needed. While such a rule is easy to implement, it ignores the curvature of the FD and the impact on travel times. A more sophisticated rule would be to select and locate the smallest number of points such that the over- or underestimation of the vehicle speed is never exceeding a user-defined percentage. This would locate more points on the FD where curvature is highest.

6. Exact non-entropic solution to the continuous-time link model

Following Equations (21)–(24), in this section we derive specific equations for $\vec{U}_i(t)$, $\vec{V}(t)$, $\vec{u}(t)$ and $\vec{v}(t)$ under the assumption of piecewise linear $U(t)$ and $V(t)$, a nonlinear FD, and simplified fanning. We refer to this simplified kinematic wave model as the multi-step shockwave model, which can be seen as extension of existing (single-step) shockwave theory.

6.1 Piecewise linear cumulative inflows and outflows

Equations (21)–(24) involve minimising over a continuous interval of flow rates $q \in [0, Q]$, which makes this link model with a nonlinear differentiable FD generally much more difficult to solve than using an FD with (piecewise) linear branches. However, as we will show in this section, we only need to consider a finite number of discrete flow rates if we assume piecewise linear cumulative inflows and outflows as input, combined with simplified fanning as introduced in Section 5, which allows finding an exact solution over continuous time.

In case the cumulative inflow $U(t)$ is piecewise linear, the same amount of information is contained in the series of traffic states denoted by the couple (t_i, u_i), $i = 1, \ldots, I$, where u_i is the flow rate during time interval $[t_i, t_{i+1}]$. In other words, t_i is the time instant at which the inflow rate changes to u_i. We refer to *i* as a segment of the cumulative inflow. To position the cumulative inflow curve, we set $t_1 = 0$ and $U(t_1) = 0$. Since the history prior to $t = 0$ is also relevant, we assume a pre-loaded network with a stationary flow rate of u_0, such that we start with $(-\infty, u_0)$. In case we start with an empty network, then $u_0 = 0$. Finally, we assume that the last inflow rate, u_I (e.g., zero flow), is stationary and lasts indefinitely, such that we set $t_{I+1} = \infty$.

Similarly, if the cumulative outflow $V(t)$ is piecewise linear, we can formulate the series (t_j, v_j), $j = 1,\ldots,J$, where v_j is the flow rate on time interval $[t_j, t_{j+1}]$. Let $t_1^* = L/\gamma_I(u_0)$ denote the time instant for which the inflow at $t = 0$ exits the link.[4]

Since the cumulative inflow and outflow is piecewise linear, we can write

$$U(t) = U(t_i) + (t - t_i)u_i, \quad \text{if } t_i \le t \le t_{i+1}, \text{ for } i = 1,\ldots,I, \tag{31}$$

$$V(t) = V(t_j) + (t - t_j)v_j, \quad \text{if } t_j \le t \le t_{j+1}, \quad \text{for } j = 1,\ldots,J, \tag{32}$$

where $U(t_i)$, for all $i = 2,\ldots,I$, and $V(t_j)$, for all $j = 2,\ldots,J$, can be computed recursively, namely $U(t_i) = U(t_{i-1}) + (t_i - t_{i-1})u_{i-1}$ and $V(t_j) = V(t_{j-1}) + (t_j - t_{j-1})v_{j-1}$, respectively. For all $t < 0$ it holds that $U(t) = tu_0$ and $V(t) = (t - t_1^*)v_0$ (note that it is not a problem if absolute cumulative inflows and outflows become negative, only relative increases in cumulative flow are relevant).

6.2 Simplified model for hypocritical states that move downstream

We first look at the case when the upstream link boundary is in a hypocritical state, so therefore the link is uncongested or congested, but not in spillback. We can calculate $\tilde{U}(t)$ in Equation (21) by looking at the range of influence[5] of each segment *i*. A segment *i* with flow rate u_i can only influence the potential outflow rate during a limited time interval as illustrated by the red lines in Figure 10(a). Let $\mathcal{T}_i$ denote the time interval of influence of segment *i*, which can be written as

$$\mathcal{T}_i = \left\{ t \in [0,T] \,|\, t_i + \frac{L}{\gamma_I(u_i)} \le t \le t_{i+1} + \frac{L}{\gamma_I(u_i)} \right\}. \tag{33}$$

Let $\mathcal{I}(t) = \{ i \in \{0,\ldots,I\} \,|\, t \in \mathcal{T}_i \}$ denote the set of segments of the cumulative inflow that can impact the sending flow rates at time instant *t*, and can be seen as the domain of dependence in terms of segments. As a result, $\mathcal{U}(t) = \{ u_i \,|\, i \in \mathcal{I}(t) \}$ is the set of discrete inflow rates that needs to be considered in the minimisation in Equations (21) and (22).

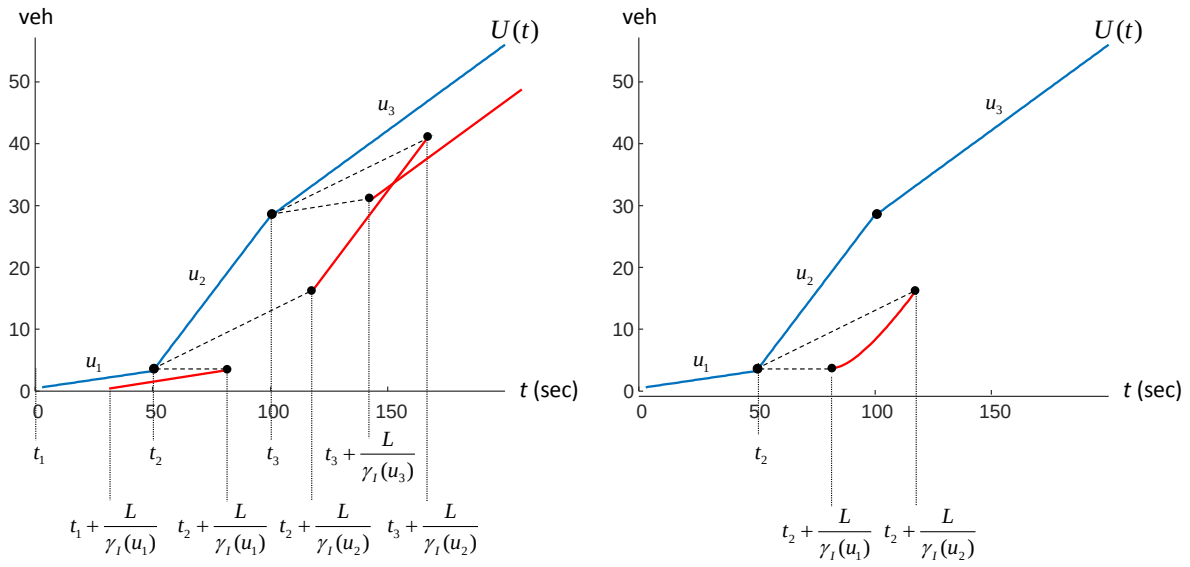
As shown in Figure 10(a), each segment has its own time interval of influence, which may overlap. What can also be observed is that no segments have an influence during the fanning time period. Hence, we also need to consider other flow rates that may be valid in case of fanning as illustrated by the red line in Figure 10(b). If there is a flow rate increase at time instant t_i, then this can have an impact on the sending flow rate during the fanning time period in which any flow rate between u_{i-1} and u_i needs to be considered. Let $\mathcal{T}_{F,i}$ denote the time interval of influence of an increase in flow rate at time instant t_i, which is defined as

$$\mathcal{T}_{F,i} = \left\{ t \in [0,T] \,|\, t_i + \frac{L}{\gamma_I(u_{i-1})} \le t \le t_i + \frac{L}{\gamma_I(u_i)} \right\}. \tag{34}$$

Let $\mathcal{I}_F(t) = \{ i \in \{0,\ldots,I\} \,|\, t \in \mathcal{T}_{F,i} \}$ denote the set of segments that constitute an increase in inflow rate. Then $\mathcal{U}_F(t) = \{ q \in [0,Q] \,|\, u_{i-1} \le q \le u_i, i \in \mathcal{I}_F(t) \}$ is the set of continuous intervals of fanning flow rates to be considered. We note that in case the flow decreases, then it holds that $\gamma_I(u_i) < \gamma_I(u_{i-1})$ such that by definition $\mathcal{T}_{F,i}(t) = \varnothing$ and hence $\mathcal{I}_F(t) = \varnothing$. In other words, this definition in compatible with the notion that fanning effects only occur with flow increases.

[4] Note that for notational convenience we are using the same variable names $t_0, t_1, \ldots$ etc. to denote the time instant that the inflow rate changes as well as the time instant that the outflow rate changes. Clearly, in general these are different time instants. We believe this should not cause confusion in the remainder of this section as they will never appear together in the same equation, and we will use subindex *i* when we refer to changes in inflow and to subindex *j* when we refer to changes in outflow.

[5] Daganzo (2005) uses the terms 'range of influence' and 'domain of dependence' to indicate both space and time intervals that indicate the relationship between actual inflows (outflows) and potential outflows (inflows).



(a) Regular influence of inflow rates (b) Fanning influence of increasing inflow rates

Figure 10: Time intervals of influence of different inflow rates

Going back to Example 1, denoting time instants in seconds (rounded off), the segments of the cumulative inflows are defined as $(t_0, u_0) = (-\infty, 200)$, $(t_1, u_1) = (0, 200)$, $(t_2, u_2) = (50, 1800)$, and $(t_3, u_3) = (100, 1000)$. This results in downstream time periods of influence $\mathcal{T}_0 = (-\infty, 31]$, $\mathcal{T}_1 = [31, 81]$, $\mathcal{T}_2 = [117, 167]$, $\mathcal{T}_3 = [140, \infty)$, and $\mathcal{T}_{F,2} = [81, 117]$, as shown in Figure 10. Let us now look at the content of sets $\mathcal{I}(t)$, $\mathcal{I}_F(t)$, and $\mathcal{U}(t)$, and $\mathcal{U}_F(t)$ for this same example. Table 2 holds some representative time instances and the values that these sets take on accordingly.

Table 2: Example 1 reference times and (non-)fanning set contents

Time t (s)	*Non-fanning*		*Fanning*	
	$\mathcal{I}(t)$	$\mathcal{U}(t)$	$\mathcal{I}_F(t)$	$\mathcal{U}_F(t)$
25	{0}	{200}	∅	∅
50	{1}	{200}	∅	∅
100	∅	∅	{2}	[200,1800]
130	{2}	{1800}	∅	∅
150	{2,3}	{1800,1000}	∅	∅
200	{3}	{1000}	∅	∅

Using sets $\mathcal{I}(t)$, $\mathcal{I}_F(t)$, $\mathcal{U}(t)$, and $\mathcal{U}_F(t)$, we can rewrite Equation (21) as

$$\vec{U}(t) = \min\left\{\min_{q\in\mathcal{U}(t)}\left\{U\left(t-\frac{L}{\gamma_I(q)}\right)+\xi_I(q)\right\}, \min_{q\in\mathcal{U}_F(t)}\left\{U\left(t-\frac{L}{\gamma_I(q)}\right)+\xi_I(q)\right\}\right\} \tag{35}$$

$$= \min\left\{\min_{i\in\mathcal{I}(t)}\vec{U}_i(t), \min_{i\in\mathcal{I}_F(t)}\vec{F}_i(t)\right\} \tag{36}$$

$$\approx \min\left\{\min_{i\in\mathcal{I}(t)}\vec{U}_i(t), \min_{i\in\mathcal{I}_F(t)}\vec{F}_i^{\,X}(t\,|\,\cdot)\right\}. \tag{37}$$

In other words, in Equation (35) we split the minimisation problem into two sub-problems, one describing the regular influence of inflow rates, and one describing the fanning influence of inflow increases. In Equation (36) we define projected linear line segments $\vec{U}_i(t)$ that result from linear

cumulative inflow segments, and we define projected nonlinear forward fanning line segments $\vec{F}_i(t)$. Finally, in Equation (37) we introduce the simplified fanning mechanism in the general formulation which approximates $\vec{F}_i(t)$ by piecewise linear forward fanning line segments $\vec{F}_i^X(t\,|\,\cdot)$, where *X* is either 'in' to represent inner linearization, or 'out' to represent outer linearization. D_i-step inner linearization is represented by $\vec{F}_i^{\text{in}}(t\,|\,u_{i-1}, q_{i,d}, \ldots, q_{i,D_i-1}, u_i)$ with intermediate flow rates $q_{i,d}$, $d = 1, \ldots, D_i - 1$ (where we explicitly note that the number of steps, D_i, can vary across segments *i* according to a pre-defined rule, see Section 6.4), and for D_i-step outer linearization we use piecewise linear line segments $\vec{F}_i^{\text{out}}(t\,|\,u_{i-1}, q_{i,2}, \ldots, q_{i,D_i-1}, u_i)$.

While Equation (36) still points to an entropic solution, the solution in Equation (37) does in general not achieve an exact minimum and hence results in a non-entropic solution. Again, we point out that the simplification only affects the fanning part, while the other parts remain unchanged. In other words, this formulation only temporarily differs from the original formulation by approximating it, before returning to the same solution once we leave the period of influence of the fanning part.

Projected linear line segments $\vec{U}_i(t)$ can be obtained by substituting Equation (31) into Equation (21), such that for each segment $i \in \mathcal{I}(t)$,

$$\vec{U}_i(t) = U(t_i) + \left(t - t_i - \frac{L}{\gamma_I(u_i)} \right) u_i + \xi_I(u_i), \quad \text{if } t \in \mathcal{T}_i. \tag{38}$$

We assume that $\vec{U}_i(t) = \infty$ for all $t \notin \mathcal{T}_i$. Without repeating this, we assume throughout this section that the value of any (fanning or non fanning) line segment is set to infinity outside the time interval of influence. The line segments in Equation (38) are indicated in red in Figure 10(a) and are translations of each segment *i* of cumulative inflow $U(t)$ that is moved $L/\gamma_I(u_i)$ to the right (i.e., forward in time) and moved $\xi_I(u_i)$ up (i.e., increased in cumulative flow).

Applying D_i-step inner linearization yields a projected piecewise linear line segment $\vec{F}_i^{\text{in}}(t\,|\,\cdot)$, which for each $i \in \mathcal{I}_F(t)$ consists of the following $D_i + 1$ adjacent linear line segments *d*, $d = 0, \ldots, D_i$,

$$\vec{F}_{i,d}^{\text{in}}(t\,|\,q_{i,0}, \ldots, q_{i,D_i}) = U(t_i) + \left(t - t_i - \frac{L}{\gamma_I(q_{i,d})} \right) q_{i,d} + \xi_I(q_{i,d}), \quad \text{if } t \in \mathcal{T}_{F,i,d}^{\text{in}}, \tag{39}$$

where the first and last flow rates are set to $q_{i,0} = u_{i-1}$ and $q_{i,D_i} = u_i$. For the first and last segments it holds that $\mathcal{T}_{F,i,0}^{\text{in}} = [t_i + L/\gamma_I(q_{i,0}), t_i + L/\eta_I(q_{i,0}, q_{i,1})]$ and $\mathcal{T}_{F,i,D_i}^{\text{in}} = [t_i + L/\eta_I(q_{i,D_i-1}, q_{i,D_i}), t_i + L/\gamma_I(q_{i,D_i})]$, while for intermediate segments $\mathcal{T}_{F,i,d}^{\text{in}} = [t_i + L/\eta_I(q_{i,d-1}, q_{i,d}), t_i + L/\eta_I(q_{i,d}, q_{i,d+1})]$, where $\eta_I(\cdot)$ is the forward shockwave speed defined in Equation (28). This linearization reduces the set of inflow rates that needs checking during fanning to the following set: $\mathcal{U}_F^{\text{in}}(t) = \left\{ u_{i-1}, q_{i,1}, \ldots, q_{i,D_i-1}, u_i \,|\, i \in \mathcal{I}_F(t) \right\}$. Therefore, in addition to existing flow rates u_{i-1} and u_i, D_i-step inner linearization creates $D_i - 1$ new flows rates $q_{i,1}, \ldots, q_{i,D_i-1}$.

In case of D_i-step outer linearization in the hypocritical branch we would get different projected piecewise linear fanning line segments $\vec{F}_i^{\text{out}}(t\,|\,\cdot)$, which can be written for each $i \in \mathcal{I}_F(t)$ as the following series of $D_i - 1$ adjacent linear line segments *d*, $d = 1, \ldots, D_i - 1$,

$$\vec{F}_{i,d}^{\text{out}}(t\,|\,q_{i,1}, \ldots, q_{i,D_i}) = U(t_i) + \left(t - t_i - \frac{L}{\gamma_I\left(\phi_I(q_{i,d}, q_{i,d+1})\right)} \right) \phi_I(q_{i,d}, q_{i,d+1}) + \xi_I\left(\phi_I(q_{i,d}, q_{i,d+1})\right), \text{ if } t \in \mathcal{T}_{F,i,d}^{\text{out}}, \tag{40}$$

where again the first and last flow rates are set to $q_{i,1} = u_{i-1}$ and $q_{i,D_i} = u_i$, where intermediate flow rates $\phi_I(\cdot)$ are defined in Equation (30), and $\mathcal{T}_{F,i,d}^{\text{out}} = [t_i + L/\gamma_I(q_{i,d}), t_i + L/\gamma_I(q_{i,d+1})]$. The set of inflow rates that need checking during fanning reduces to the following set: $\mathcal{U}_F^{\text{out}}(t) = \left\{ \phi_I(u_{i-1}, q_{i,2}), \phi_I(q_{i,2}, q_{i,3}), \ldots, \phi_I(q_{i,D_i-1}, u_i) \,|\, i \in \mathcal{I}_F(t) \right\}$. Therefore, D_i-step outer linearization creates $D_i - 1$ new flows rates while using none of the existing flow rates.

The fact whether new flow rates are created or not becomes important when we discuss computational efficiency in Section 7. As is clear from Figure 9, 1-step inner linearization does not create any new flow rates, while 2-step outer linearization creates one new flow rate.

In order to determine the potential outflow rate $\vec{u}(t)$ as defined in Equation (22) we need to find the flow rate that belongs to the line segment that minimises Equation (37). For example, if we use 2-step outer linearization, then we need to check flow rates $\mathcal{U}(t)$ and $\mathcal{U}_F^{\text{out}}(t)$. If for time instant *t* the minimum is achieved by $i^* \in \mathcal{I}(t)$, then the sending flow rate is equal to $\vec{u}(t) = u_{i^*}$, while if the minimum is achieved by $i^* \in \mathcal{I}_F(t)$, then $\vec{u}(t) = \phi_I(u_{i^*-1}, u_{i^*})$.

6.3 Link model for hypercritical traffic states that move upstream

We now look at the case when the downstream link boundary is in a hypercritical state, so therefore the link is congested with or without spillback. We can calculate $\overleftarrow{V}(t)$ and $\overleftarrow{v}(t)$ as established in Equation (23) and (24) in a similar way as discussed for $\vec{U}(t)$ and $\vec{u}(t)$ in the previous section.

Equation (23) can be written as

$$\overleftarrow{V}(t) = \min\left\{\min_{q\in\mathcal{V}(t)}\left\{V\left(t+\frac{L}{\gamma_{II}(q)}\right)+\xi_{II}(q)\right\}, \min_{q\in\mathcal{V}_F(t)}\left\{V\left(t+\frac{L}{\gamma_{II}(q)}\right)+\xi_{II}(q)\right\}\right\} \tag{41}$$

$$\approx \min\left\{\min_{j\in\mathcal{J}(t)}\overleftarrow{V}_j(t), \min_{j\in\mathcal{J}_F(t)}\overleftarrow{F}_j^X(t\,|\,\cdot)\right\}. \tag{42}$$

For brevity and to avoid repetitive statements, we refer to Appendix C for definitions of sets $\mathcal{V}(t)$, $\mathcal{V}_F(t)$, $\mathcal{J}(t)$, and $\mathcal{J}_F(t)$, as well as definitions of projected linear line segments $\overleftarrow{V}_j(t)$, and projected piecewise linear line segments $\overleftarrow{F}_j^X(t\,|\,\cdot)$ for inner or outer linearization. It again holds that using D_j-step (inner or outer) linearization we add $D_j - 1$ new flow rates.

6.4 Lower envelope

Equations (37) and (42) can be seen as lower envelope problems with *linear* line segments (due to our simplified fanning). These are relatively easy problems to solve; it is well-known that finding the lower envelope of *z* linear line segments has a computational complexity of $O(z \log z)$, see e.g., Hershberger (1989). Therefore, Equations (37) and (44) can be solved exactly, not only for specific time instants, but over any continuous time period.

Key to solving a lower envelope problem is to determine at what time instants different line segments intersect. In the link model for downstream hypocritical traffic states there can be multiple intersecting segments, namely an intersection of regular forward segments $\vec{U}_i(t)$, an intersection of forward fanning segments $\vec{F}_i^X(t\,|\,\cdot)$, or a combination of the two. Similarly, in the link model for hypercritical traffic states we can have an intersection of regular backward segments $\overleftarrow{V}_j(t)$, an intersection of backward fanning segments $\overleftarrow{F}_i^X(t\,|\,\cdot)$, or a combination of the two.

Consider a situation with a free-flow hypocritical state at the upstream link boundary. Then $\vec{t}_{i_1,i_2}$ is the time instant that two regular line segments $\vec{U}_{i_1}(t)$ and $\vec{U}_{i_2}(t)$ intersect (with $i_1 < i_2$) if and only if holds that (i) $\vec{t}_{i_1,i_2} \in \mathcal{T}_{i_1}$ and $\vec{t}_{i_1,i_2} \in \mathcal{T}_{i_2}$ (i.e., their time periods of influence overlap), and (ii) $\vec{U}_{i_1}(\vec{t}_{i_1,i_2}) = \vec{U}_{i_2}(\vec{t}_{i_1,i_2})$. It can be shown that $\vec{t}_{i_1,i_2}$ can be calculated as (see Appendix D)

$$\vec{t}_{i_1,i_2} = t_{i_1,i_2} + \frac{L}{\eta_I(u_{i_1}, u_{i_2})}, \quad \text{where } t_{i_1,i_2} = \frac{\sum_{i=i_1+1}^{i_2} t_i(u_i - u_{i-1})}{u_{i_2} - u_{i_1}}. \tag{43}$$

In other words, $\vec{t}_{i_1,i_2}$ can be calculated by following the shockwave between flow rates u_{i_1} and u_{i_2} to the end of the link starting at a virtual 'representative' time instant t_{i_1,i_2} that accounts for variations in the

flow rate during interval $[t_{i_1}, t_{i_2}]$. It is easy to see that if we consider two adjacent segments in which $i_2 = i_1 + 1$, then the representative time instant simplifies to t_{i_2}. Equation (43) is illustrated in Figure 11, where we consider an inflow rate that decreases from u_1 to u_2 to u_3. This initially generates two shockwaves that later merge to a single shockwave, which arrives on the opposite border (and hence is the only shockwave we are interested in). While existing front wave tracking algorithms explicitly consider interactions between shockwaves across the link, Equation (43) shows that it is not necessary to explicit track each shockwave, but rather that only information on the upstream link boundary is required. We can derive a similar equation for the intersection of line segments $\vec{V}_{j_1}(t)$ and $\vec{V}_{j_2}(t)$.

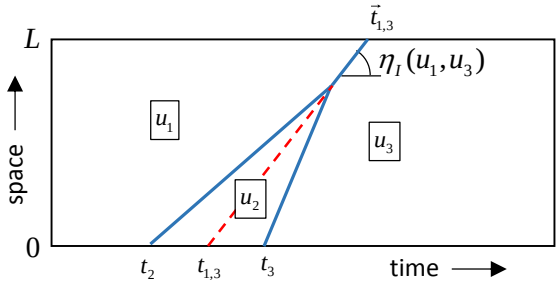


Figure 11: Implicitly determining flow rate changes at the opposite border

Now we look at piecewise linear fanning line segments $\vec{F}_i^X(t|\cdot)$. First, consider the case of inner linearization. Suppose there are no further intersections with this line segment. In that case the lower envelope would simplify follow the linear sub-segments $\vec{F}_{i,d}^{\text{in}}(t|\cdot)$, which according to the definition of $\mathcal{T}_{F,i,d}^{\text{in}}$ provided in Section 6.2 connect at time instants that can be determined by following shockwaves. If multiple fanning events at time instants t_{i_1} and t_{i_2} interact with each other, then we need to determine intersections between $\vec{F}_{i_1}^{\text{in}}(t|\cdot)$ and $\vec{F}_{i_2}^{\text{in}}(t|\cdot)$. Given the fact that the structure of Equation (38) is the same as the structure of Equation (39), in case of inner linearization these intersections can again be determined by tracking shockwaves. Moreover, intersections between $\vec{U}_{i_1}(t)$ and $\vec{F}_{i_2}^{\text{in}}(t|\cdot)$ also can be determined using shockwaves. This leads to the important observation that with inner linearization we can determine an exact solution to the lower envelope problem by merely calculating shockwave speeds between prevailing inflow rates.

In contrast, when applying outer linearization, we need to track waves during fanning events instead of shockwaves as evidenced from the definitions of $\mathcal{T}_{F,i,d}^{\text{out}}$. This also means that when $\vec{U}_{i_1}(t)$ interacts with $\vec{F}_{i_2}^{\text{out}}(t|\cdot)$ we need to track combinations of waves and shockwaves. The model of Henn (2003) is consistent with our lower envelope problem assuming outer linearization. He solves this problem by tracking waves over space and time, which requires more book-keeping than simply considering shockwaves at the link boundaries only as can be done using inner linearization. Therefore, while both inner and outer linearization techniques can be used to find an exact non-entropic solution to the kinematic wave model, we believe that assuming inner linearization makes it easier to develop efficient solution algorithms.

7. Potential for event-based algorithms on networks

In Section 6 we formulated link models with simplified fanning that can be solved exactly in continuous time. In this section we discuss the potential of using an event-based algorithm to solve the link model in the context of general networks. To simplify our discussion, we consider single-commodity network flow.

7.1 Events

Following our definitions in Section 6, we let (t_i, u_i) denote an event in which the inflow changes at time instant t_i to flow rate u_i. Similarly, we let (t_j, v_j) denote an event in which the outflow changes.

Inflow and outflow rates are determined by the node model in Equation (13). It is clear from this equation that in a single-commodity model (in which split proportions $\varphi_{ab}(t)$ are considered stationary input) inflow and outflow rates can only change when either sending flow rates $s(t)$ or receiving flow rates $r(t)$ change.

A sending flow rate as defined in Equation (9) only changes if potential outflow rate $\vec{u}(t)$ changes (via the link model in Equation (22)), or if there is a change in the traffic state from free-flow to congestion, or vice versa. A receiving flow rate as defined in Equation (10) only changes if the potential inflow rate $\vec{v}(t)$ changes (via the link model in Equation (24)), or if there is a change in the traffic state from spillback to congestion, or vice versa. Each of these changes would create an event. In Section 6.4 we already discussed that events in the link model (i.e., changes in $\vec{u}(t)$ and $\vec{v}(t)$) can be determined by solving a lower envelope problem in which we need to determine intersections between linear line segments by tracking projected shockwaves and/or characteristic waves. In order to determine when link states change, we need to determine intersections between the cumulative potential outflow $\vec{U}(t)$ and the actual cumulative outflow $V(t)$ to track transitions between free-flow and congested states, while we need to find intersections between cumulative potential inflows $\vec{V}(t)$ and actual cumulative inflow $U(t)$ for changes between congested and spillback states. These state changes can again be found by implicitly tracking shockwaves.

In the next sections we discuss the potential of using event-based algorithms to apply LTM-based DNL on general networks by making a comparison with the efficient event-based algorithm described in Raadsen et al. (2016) – which we believe is currently one of the most efficient algorithms that is capable of finding exact or approximate solutions in continuous time – that relies on a triangular FD as input. We refer to Raadsen et al. (2016) for a comparison between time-discretised algorithms and continuous-time (event-based) algorithms.

7.2 Computational efficiency on a single link

The efficiency of any event-based algorithm heavily depends on the number of segments in the cumulative inflow and outflow curves that are input into the link model, no matter what FD is considered.

What makes using a smooth nonlinear FD more complex is the necessity to minimise over multiple flow rates as indicated in Equations (21) and (23) compared to Equations (25) and (26) for a triangular FD. Therefore, in order to calculate potential inflow and outflow rates, multiple segments of the cumulative curves need to be evaluated. Solving the link model for a non-entropic solution as in Equations (37) and (42) will therefore lead to an increase in computation time. However, we note that the time periods of influence of each segment are limited and in many cases there will exist only few overlapping line segments (e.g., in Figure 10 only during a short time period we need to consider two flow rates instead of one). Further, if the number of steps D in the linearization is limited, then again only a small number of flow rates need to be evaluated in the minimisation. Therefore, we expect that in practice the computational efficiency when using smooth nonlinear FDs will only be slight higher compared to using a triangular FD; the computational burden will only be significantly affected in case the flow rates change with a very high frequency, and if D is high.

7.3 Computational efficiency on a network

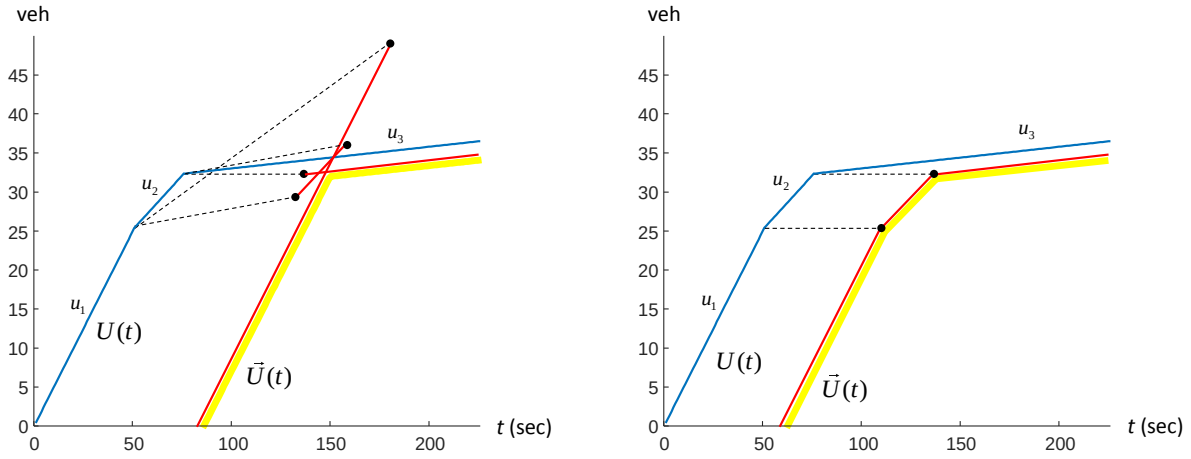
While computational efficiency of a single link is predominantly a function of the number of events as input, for the computational efficiency on a network also the number of events as output are relevant since they result in new events for adjacent links. In this discussion the shape of the FD and the approximation of the FD using linearization both play an important role. First, we discuss the influence of on-the-fly D-step inner or outer linearization of the FD, and then we make comparisons with linear and piecewise linear FDs.

Events can constitute either a flow increase or a flow decrease, and both have different impacts on the number of events generated. Consider a link that is in a free-flow state and assume a piecewise linear cumulative inflow curve with $Z+1$ segments with Z_1 flow increases and Z_2 flow decreases (where $Z_1+Z_2=Z$). From Section 6 we know that in case the link is in a free-flow or spillback state the D-step (inner or outer) linearization creates $D-1$ additional intermediate flow rates. As a result, applying D-step linearization to all flow increases would result in a potential cumulative outflow curve that has Z_1D (smaller) flow increases. In case the link remains in a free-flow state, this will lead to $Z_1(D-1)$ *additional* events that are passed on to downstream links. If we consider a corridor network with Y consecutive links that are all in free-flow, then the Z_2 flow increase events into the first link will result in Z_1D^Y outflow increase events out of the last link if we apply D-step linearization. In other words, the number of events can easily grow exponentially if $D>1$.

The ability to choose $D=1$ in the inner linearization technique is clearly an advantage, since this means that the number of events does not increase due to expansion fans. Outer linearization in which $D\geq 2$ will significantly increase the number of events. With respect to inflow rate decreases, the Z_2 events will result in *at maximum* Z_2 changes in the potential outflow rate (which is illustrated in Example 2 below).

The exact same reasoning can be held for a link that is in a spillback state in which flow increases may lead to an explosion of network events, while flow decreases are of less concern. Note that for links that are in a congested state, the sending and receiving flows are not affected by changes in the inflow and outflow rates, and as such congested links act as filters that remove events. This leads to the interesting observation that an event-based algorithm is likely more efficient on highly congested networks than on free-flow networks, even if many links are in spillback (since backward wave speeds are slower than forward wave speeds).

Example 2 – Consider a link with $L=2$ (km), $Q=2000$ (veh/h). $\gamma^{\max}=120$ (km/h), $\sigma^{\text{crit}}=80$ (km/h), and $K=180$ (veh/km), and assume a fundamental diagram with a quadratic hypocritical branch. Assume that the link inflow rate is stationary at 1800 veh/h for some time, but that at $t=50$ s the inflow rate decreases to 1000 veh/h, and drops to 100 veh/h at $t=75$ s. Figure 12(a) shows the cumulative inflow as well as the potential cumulative outflow as the lower envelope (in yellow). While there are three inflow rates, there are only two potential outflow rates. Inflow rates of 1000 veh/h are not able to influence the potential outflow rate, since the flow rate of 100 veh/h has a dominating influence in the lower envelope. This situation coincides with the illustration in Figure 11. For comparison, we also calculate the result when using a triangular FD in which $\sigma^{\text{crit}}=\gamma^{\max}$, see Figure 12(b). In this case, potential outflow rates simplify mirror inflow rates at a later time instant. ■



(a) Quadratic hypocritical branch (b) Linear hypocritical branch

Figure 12: Potential cumulative outflows considering different shapes of the fundamental diagram

Now let us compare the number of events generated under different shapes of the FD. As shown in Figure 12(b), if we would have used a triangular FD in Example 1 this would have led to three potential outflow events, instead of the two in Figure 12(a). This means that while the computations with a smooth nonlinear FD for a single link as discussed in Section 7.2 are slightly more complex, the total number of events propagating through the network can actually be smaller (particularly in case of 1-step inner linearization).

With respect to piecewise linear FDs, the number of segments in a hypocritical or hypercritical branch works similar to the number of steps used in our linearization. Consider a link in free-flow or spillback with Z_1 flow increase events as input and assume a piecewise linear branch in the FD with *S* segments. Then the number of generated events will at best be Z_1 (if both flows are on the same segment) and at worst $Z_1(S+1)$ (if the flow rates belong to the first segment and last segments, respectively). As such, using a piecewise linear branch has a similar computational efficiency as using a smooth nonlinear branch with a *D*-step linearization where $1 \le D \le S+1$. Therefore, we expect for example that the computational efficiency with two branches in a piecewise linear FD is on average comparable with 2-step inner or outer linearization. That said, in practice it is likely that the number of steps is not a fixed number but varies during the simulation as discussed earlier. In that case, on-the-fly inner linearization will then only introduce more events when necessary while only using a single step in other cases. A piecewise linear FD on the other hand is much more rigid. One can have a large change in the flow rate within one of the linear segments resulting in no additional fanning steps, while a small flow rate change may be across two segments resulting in multi-step fanning, which is undesirable.

8. Summary and discussion

In this paper we have derived continuous-time LTM formulations assuming an FD with smooth nonlinear branches. Using smooth nonlinear branches has several advantages in terms of realism, traffic equilibrium properties, and ease of calibration. The main challenge has been to develop model formulations that can be almost as efficiently solved as simplified triangular or piecewise linear FDs as input. To this end we have proposed on-the-fly linearization techniques of the FD within the model formulation itself to simplify expansion fans, which allows for determining exact solutions and also open up the possibility of developing event-based algorithms that can be applied to conduct dynamic network loading on general (large scale) networks.

The simplest LTM based formulation with a smooth nonlinear FD adopts 1-step inner linearization consistent with shockwave theory. We argued that this model can be efficiently solved using an event-based algorithm similar to Raadsen et al. (2016), in which the number of events generated will be in close proximity to the number of events needed when one adopts a triangular fundamental diagram. While the calculation time for each event is expected to slightly increase, calculation time will mainly increase if the number of events increases. Using multi-step linearization will provide better approximation of the expansion fans, but will likely come at a significantly higher cost due to the increase in the number of generated events in the network.

The paper has several limitations. First of all, we have only provided theory for a single link and did not formulate any specific solution algorithms. Second, our arguments have been based on theory and we did not support our claims regarding computational efficiency on networks with actual calculations. Thirdly, we assumed single-commodity flow, thereby ignoring route choice behaviour in DNL. These three aspects will be addressed in Part II of this study (see Raadsen and Bliemer, 2017).

Further, we ignored traffic controls (such as traffic lights at intersections). Such controls can be seen as yet another type of event that is either activated by a fixed control scheme, or can be traffic actuated. Therefore, we believe that it is possible to add such traffic controls within our model framework. Relaxation to include multiple vehicle types is more difficult. LTM provides an efficient approach to

solving the kinematic wave model under the assumption of first-in-first-out (FIFO). As far as we are aware, there do not yet exist multiclass extensions to LTM that can be efficiently solved.

References

Aubin, J.P., A.M. Bayen, and P. Saint-Pierre (2008) Dirichlet problems for some Hamilton-Jacobi equations with inequality constraints. *SIAM Journal on Control and Optimization*, Vol. 47(5), pp. 2348-2380.

Boyles, S.D., J. Duthie, C. Melson, and T. Rambha (2013) Diverge models and dynamic traffic equilibria. *Presented at the Annual Meeting of the Institute for Operations Research and Management Sciences* (INFORMS), Minneapolis, MN, USA.

Bliemer, M.C.J., M.P.H. Raadsen, and M.G.H. Bell (2016) Discrete and continuous time formulations of the link transmission model class. *Presented at the 6th International Symposium on Dynamic Traffic Assignment*, Sydney, Australia.

Cascetta, E. (2009) *Transportation Systems analysis: model and applications*. Second edition. Springer

Claudel, C.G. and Bayen, A.M. (2010a) Lax–Hopf Based Incorporation of Internal Boundary Conditions Into Hamilton–Jacobi Equation. Part I: Theory. *IEEE Transactions on Automatic Control*, 55(5), pp. 1142–1157.

Claudel, C.G. and Bayen, A.M. (2010b) Lax–Hopf Based Incorporation of Internal Boundary Conditions Into Hamilton–Jacobi Equation. Part II: Computational Methods. *IEEE Transactions on Automatic Control*, 55(5), pp. 1158–1174.

Daganzo, C.F. (1994) The cell transmission model: A dynamic representation of highway traffic consistent with the hydrodynamic theory. *Transportation Research Part B*, Vol. 28(4), pp. 269-287.

Daganzo, C.F. (1995) The cell transmission model, part II: Network traffic. *Transportation Research Part B*, Vol. 29(2), pp. 79-93.

Daganzo, C.F. (2001) A simple traffic analysis procedure. *Networks and Spatial Economics*, Vol. 1, pp. 77-101.

Evans, L.C. (1998) *Partial differential equations*. American Mathematical Society, Providence, RI, USA.

Friesz T.L., Han K., and Neto P. (2013) Dynamic user equilibrium based on a hydrodynamic model. *Transportation Research Part B*, Vol. 47, pp.102–126.

Gentile G. (2010) The general link transmission model for dynamic network loading and a comparison with the DUE algorithm. In Tampère, C.M.J., F. Viti, and L. Immers (eds) New Developments in Transport Planning: Advances in Dynamic Traffic Assignment. Edward Elgar, Northampton MA, USA, pp. 153-178.

Greenshields, B.D. (1935) A study of highway capacity. *Proceedings Highway Research Record*, Vol. 14., pp. 448-477.

Han, J., B. Piccoli, and T.L. Friesz (2016a) Continuity of the path delay operator for dynamic network loading with spillback. *Transportation Research Part B*, Vol. 92, pp. 211-233.

Han, K., B. Piccoli, and W.Y. Szeto (2016b) Continuous-time link-based kinematic wave model: formulation, solution existence, and well-posedness. *Transportmetrica B: Transport Dynamics*, Vol. 4(3), pp. 187-222.

Hershberger, J. (1989) Finding the upper envelope of n line segments in O(n log n) time. *Information Processing Letters*, Vol. 33(4), pp. 169-174.

Henn, V. (2003) A wave-based resolution scheme for the hydrodynamic LWR traffic flow model. In: S.P. Hoogendoorn & B.S.W. Luding (eds) *Proceedings of the Workshop in Traffic and Granular Flow*, Amsterdam, pp. 105-124.

Himpe, W., R. Corthout, and C.M.J. Tampère (2016) An efficient iterative link transmission model. *Transportation Research Part B*, Vol. 92, pp. 170-190.

Jin, W-L. (2015) Continuous formulations and analytical properties of the link transmission model. *Transportation Research Part B*, Vol 74, pp. 88-103.

Lebacque, P. (1996) The Godunov scheme and what it means for first order traffic flow models. In J. B. Lesort (ed.) *Proceedings of the 13th ISTTT Symposium*, Elsevier, New York, pp. 647-678.

Lebacque, J., Khoshyaran, M. (2005) First order macroscopic traffic flow models: intersections modeling, network modeling. In: Mahmassani (ed) *Flow, dynamics, and human interaction.* Proceedings of the 16th International Symposium of Transportation and Traffic Theory (ISTTT), Elsevier Science, Amsterdam.

Lighthill, M.J., and G.B. Whittam (1955) On kinematic waves: II. A theory of traffic flow on long crowded roads. *Proceedings of the Royal Society*, London, Ser. A 229(1178), pp. 317-345.

Lu, Y., S.C. Wong, M. Zhang, C.-W. Shu, and W. Chen (2008) Explicit construction of entropy solutions for the Lighthill-Whitham-Richards traffic flow model with a piecewise quadratic flow-density relationship. *Transportation Research Part B*, Vol. 42, pp. 355-372.

Luke, J.C. (1972) *Mathematical models for landform evolution. Journal of Geophysical Research*, Vol. 77(14), pp. 2460-2464.

May, A.D. (1990) *Traffic flow fundamentals*. Prentice-Hall, Inc., Englewood Cliffs, NJ, USA.

Mazaré, P.E., Dehwah A.H., Claudel, C.G., Bayen, A.M. (2011) Analytical and grid-free solutions to the Lighthill–Whitham–Richards traffic flow model. *Transportation Research Part B*, Vol. 45 (10), pp. 1727–1748.

Newell, G.F. (1993a) A simplified theory of kinematic waves in highway traffic, Part I: General theory. *Transportation Research Part B*, Vol. 27(4), pp. 281-287.

Newell, G.F. (1993b) A simplified theory of kinematic waves in highway traffic, Part II: Queueing at freeway bottlenecks. *Transportation Research Part B*, Vol. 27(4), pp. 281-303.

Newell, G.F. (1993c) A simplified theory of kinematic waves in highway traffic, Part III: Multi-destination flows. *Transportation Research Part B*, Vol. 27(4), pp. 305-313.

Raadsen, M.P.H., and M.C.J. Bliemer (2017) Continuous-time general link transmission model with simplified fanning, Part II: event-based algorithm for networks. Submitted for publication to Transportation Research Part B.

Raadsen, M.P.H., M.C.J. Bliemer, and M.G.H. Bell (2016) An efficient and exact event-based algorithm for solving simplified first order dynamic network loading problems. *Transportation Research Part B*, Vol. 92, pp. 191-210.

Richards, P.I. (1956) Shock waves on the highway. *Operations Research*. Vol. 4, 42-51.

Smits, E.S., M.C.J. Bliemer, A.J. Pel, and B. van Arem (2015) A family of macroscopic node models. *Transportation Research Part B*, Vol. 74, pp. 20-39.

Smulders, S. (1988) Modelling and filtering of freeway traffic flow. *Report OS-R8706*, Centre of Mathematics and Computer Science, the Netherlands.

Tampère, C.M.J. , Corthout, R., D. Cattrysse and L.H. Immers (2011) A generic class of first order node models for dynamic macroscopic simulation of traffic flows, *Transportation Research Part B*, Vol. 45, pp. 289-309.

Wong, S.C., and G.C.K. Wong (2002) An analytical shock-fitting algorithm for LWR kinematic wave model embedded with linear speed-density relationship. *Transportation Research Part B*, Vol. 36(8), pp. 683-706.

Yperman, I. (2007) *The link transmission model for dynamic network loading*. PhD Thesis, Katholieke Universiteit Leuven, Belgium.

Yperman, I., S. Logghe, and L. Immers (2005) The link transmission model: an efficient implementation of the kinematic wave theory in traffic networks. *Advanced OR and AI methods in transportation*, Poznan University of Technology, pp. 122-127.

Appendix A: Kinematic wave model and Lax-Hopf solution

A.1 Kinematic wave model

Let $k(x,t) \in [0,K]$ (veh/km) and $q(x,t) \in [0,Q]$ (veh/h) respectively denote the density and flow at location $x \in [0,L]$ at time instant $t \in [0,T]$. Let the fundamental relationship between flow and density be giving by

$$q(x,t) = \Phi\big(k(x,t)\big), \tag{A.1}$$

where $\Phi : [0,K]$ (veh/km) $\rightarrow [0,Q]$ (veh/h) is a given flux function (see Section 2). Note that in general this function does not need to be concave or differentiable. The well-known first order kinematic wave model can be written as the following partial differential equation (Lighthill and Whitham, 1955; Richards, 1956):

$$\frac{\partial k(x,t)}{\partial t} + \frac{\partial \Phi\big(k(x,t)\big)}{\partial x} = 0. \tag{A.2}$$

An alternative formulation of the kinematic wave model is given by Newell (1993a) using a Moskowitz function $N(x,t)$ (veh) that describes the cumulative number of vehicles that have passed location x up till time instant t. Since $k(x,t) = -\partial N(x,t)/\partial x$ and $q(x,t) = \partial N(x,t)/\partial t$, the fundamental relationship in Equation (A.1) can be rewritten into the following Hamilton-Jacobi equation:

$$\frac{\partial N(x,t)}{\partial t} - \Phi\left(-\frac{\partial N(x,t)}{\partial x}\right) = 0. \tag{A.3}$$

A.2 Lax-Hopf formula

In case flux function Φ is concave, the semi-analytical solution to the Hamilton-Jacobi equation (A.3) is given by the Lax-Hopf formula (Evans, 1998),

$$N(x,t) = \inf_{(x',t')} \left\{ N(t',x') + (t-t')\Phi^*\left(\frac{x-x'}{t-t'}\right)\right\}, \tag{A.4}$$

where $\Phi^* : [\gamma^{\min}, \gamma^{\max}]$ (km/h) $\rightarrow [0, -\gamma^{\min} K]$ (veh/h) is the convex conjugate of Φ. The infimum in Equation (A.4) is over all $x' \in [0,L]$ and $t' < t$ within the domain of Φ^*. Function Φ^* is also referred to as the Legendre-Fenchel transform, and can be seen as a formulation of the same function in dual space using wave speed *w* as a conjugate variable,

$$\Phi^*(w) = \sup_{k \in [0,K]} \big\{\Phi(k) - wk\big\}. \tag{A.5}$$

This transform is illustrated in Figure A.1, in which Φ is given by a two-regime concave FD that is not differentiable at critical density k^{crit}. The resulting function Φ^* is a convex decreasing function of wave speed *w*. For example, $\Phi^*(w_1)$ is equal to $\Phi(k_1) - w_1 k_1$, where k_1 is the density at which slope w_1 is touching the FD (i.e., the tangent), since this density maximises the vertical distance between the FD and the line $w_1 k$. Similarly, $\Phi^*(w_2) = \Phi(k_2) - w_2 k_2$.

Substituting the definition of vehicle speed as stated in Equation (4) into Equation (A.5) yields

$$\Phi^*(w) = \sup_{k \in [0,K]} \big\{\big(\sigma(k) - w\big)k\big\}. \tag{A.6}$$

This formulation shows that $\Phi^*(w)$ essentially depends on the difference between the vehicle speed and the wave speed, see also Figure A.1(a).

The Lax-Hopf solution as stated in Equation (A.4) can be used with any concave FD and is theoretically powerful, although it is not a trivial equation to immediately understand. It expresses how upstream traffic conditions influence downstream traffic conditions (i.e., $x' > x$) assuming there are no outflow constraints, and vice versa how downstream traffic conditions influence upstream traffic conditions (i.e., $x' < x$) assuming there are no inflow constraints.

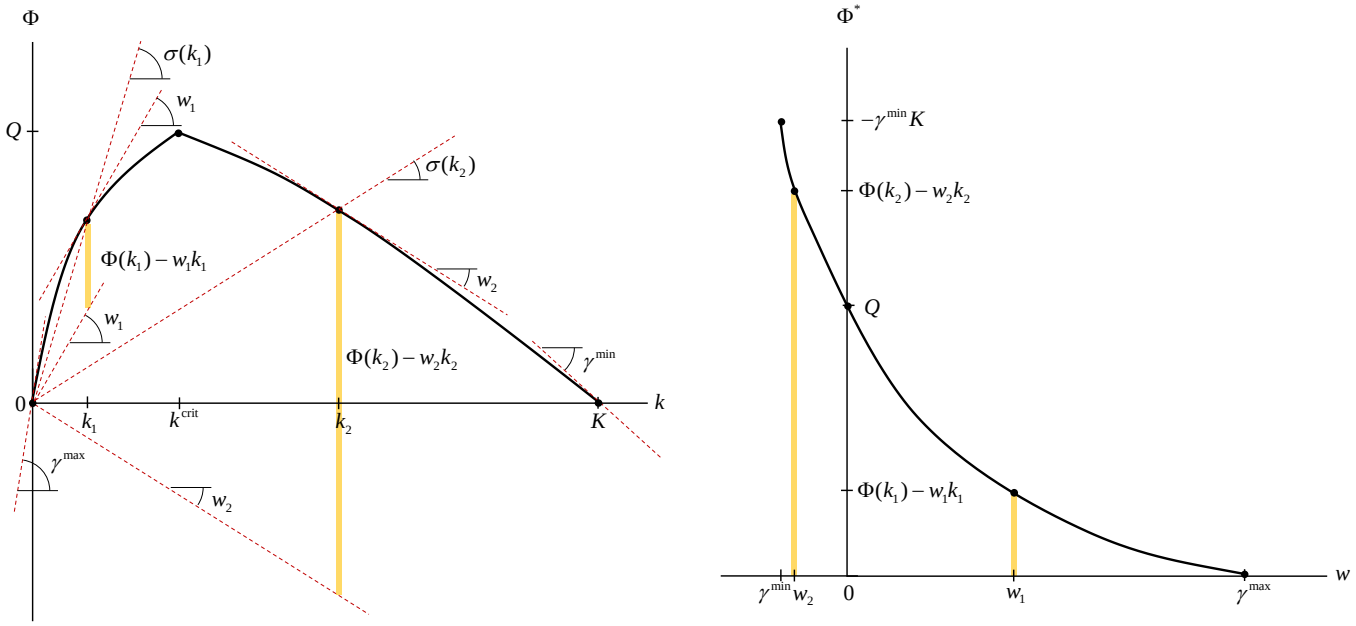


Figure A.1: Legendre-Fenchel transform of a concave fundamental diagram

Appendix B: Derivation of potential cumulative inflows

The Lax-Hopf formula can be written as

$$\overleftarrow{V}(t) = \min_{t'}\left\{V(t') + (t-t')\Phi_{II}^{*}\left(\frac{-L}{t-t'}\right)\right\}, \quad \Phi_{II}^{*}(w) = \max_{k\in[k^{\text{crit}},K]}\left\{\left(\sigma_{II}(k)-w\right)k\right\}. \tag{B.1}$$

Since $t' = t + L/w$ we can rewrite Equation (B.1) such that it minimises over (wave) speeds *w*,

$$\overleftarrow{V}(t) = \min_{w\in[\gamma^{\min},\gamma_{II}^{\text{crit}}]}\left\{V\left(t+\frac{L}{w}\right) - \frac{L}{w}\Phi_{II}^{*}(w)\right\}. \tag{B.2}$$

Since Φ_{II} is strictly decreasing, for which holds that $\gamma_{II}^{\text{crit}} < 0$, hence we can divide by *w* over the strict negative range $[\gamma^{\min}, \gamma_{II}^{\text{crit}}]$.

Define $\xi_{II}(w) = -(L/w)\Phi_{II}^{*}(w)$ (veh), which can be rewritten as

$$\xi_{II}(w) = \max_{k\in[k^{\text{crit}},K]}\left\{Lk\left(1-\frac{\sigma_{II}(k)}{w}\right)\right\}, \tag{B.3}$$

Instead of this density based formulation, it can also be written in the following equivalent flow based form,

$$\xi_{II}(w) = \max_{q\in[0,Q]}\left\{Lq\left(\frac{1}{\sigma_{II}(q)} - \frac{1}{w}\right)\right\}. \tag{B.4}$$

Hence we can write the general formula that propagates hypercritical states from the downstream to the downstream link boundary as

$$\overleftarrow{V}(t) = \min_{w\in[\gamma^{\min},\gamma_{II}^{\text{crit}}]}\left\{V\left(t+\frac{L}{w}\right) + \xi_{II}(w)\right\}. \tag{B.5}$$

It holds that $\xi_{II}(w)$ (veh) is a non-negative increasing function, with $\xi_{II}(\gamma^{\min}) = LK$. Since $V(t)$ is a monotonically increasing function, it holds that $V(t + L/w)$ is decreasing in *w*. Therefore, $V(t + L/w)$ and $\xi_{II}(w)$ work in opposite directions in the minimisation with respect to wave speed *w*.

Since function Φ_I is differentiable, the minimum in Equation (B.5) is achieved at the (unique) density $k^* \in [k^{\text{crit}}, K]$ for which holds that $w = \gamma_{II}(k^*)$. Therefore, instead of minimising over wave speeds in Equation (B.5), we can minimise over densities and simply set $w = \gamma_{II}(k)$. This results in the following density based formulation for cumulative potential inflows,

$$\overleftarrow{V}(t) = \min_{k \in [k^{\text{crit}}, K]} \left\{ V\left(t + \frac{L}{\gamma_{II}(k)} \right) + \xi_{II}(k) \right\}, \quad \text{with} \quad \xi_{II}(k) = Lk\left(1 - \frac{\sigma_{II}(k)}{\gamma_{II}(k)} \right). \tag{B.6}$$

We can also formulate a flow based version, namely

$$\overleftarrow{V}(t) = \min_{q \in [0, Q]} \left\{ V\left(t + \frac{L}{\gamma_{II}(q)} \right) + \xi_{II}(q) \right\}, \quad \text{with} \quad \xi_{II}(q) = Lq\left(\frac{1}{\sigma_{II}(q)} - \frac{1}{\gamma_{II}(q)} \right). \tag{B.7}$$

Appendix C: Link model for hypercritical traffic states that move upstream

Define the time intervals of influence of each segment *j* of the cumulative outflow curve as

$$\mathcal{T}_j = \left\{ t \in [0,T] \,|\, t_j - \frac{L}{\gamma_{II}(v_j)} \le t \le t_{j+1} - \frac{L}{\gamma_{II}(v_j)} \right\}. \tag{C.1}$$

Further, define the time interval of influence due to an increase in outflow rate at time instant t_j as

$$\mathcal{T}_{F,j} = \left\{ t \in [0,T] \,|\, t_j - \frac{L}{\gamma_{II}(v_{j-1})} \le t \le t_j - \frac{L}{\gamma_{II}(v_j)} \right\}. \tag{C.2}$$

Define the corresponding domains of dependence as $\mathcal{J}(t) = \left\{ j \in \{0,\ldots,J\} \,|\, t \in \mathcal{T}_j \right\}$ and $\mathcal{J}_F(t) = \left\{ j \in \{1,\ldots,J\} \,|\, t \in \mathcal{T}_{F,j} \right\}$, respectively. Let $\mathcal{V}(t) = \left\{ v_j \,|\, j \in \mathcal{J}(t) \right\}$ and $\mathcal{V}_F(t) = \left\{ q \in [0,Q] \,|\, v_{j-1} \le q \le v_j,\ j \in \mathcal{J}_F(t) \right\}$ be the sets of regular flow rates and fanning flow rates, respectively, that need to be considered in the minimisation in Equations (23) and (24). Then we can rewrite Equation (23) as Equation (41), which can be approximated by Equation (42) using D_j-step inner or outer linearization.

In Equation (42), $\overleftarrow{V}_j(t)$ are projected linear line segments, defined for each segment $j \in \mathcal{J}(t)$ as

$$\overleftarrow{V}_j(t) = V(t_j) + \left(t - t_j + \frac{L}{\gamma_{II}(v_j)} \right) v_i + \xi_{II}(v_j), \quad \text{if } t \in \mathcal{T}_j. \tag{C.3}$$

In case of D_j- step inner linearization we obtain the following $D_j + 1$ projected linear segments (indicated by index *d*, $d = 0,\ldots,D_j$) for each $j \in \mathcal{J}_F(t)$ that describe backwards fanning,

$$\vec{F}^{\text{in}}_{j,d}(t \,|\, q_{j,0},\ldots,q_{j,D_j}) = U(t_j) + \left(t - t_j + \frac{L}{\gamma_{II}(q_{j,d})} \right) q_{j,d} + \xi_{II}(q_{j,d}), \quad \text{if } t \in \mathcal{T}^{\text{in}}_{F,j,d}, \tag{C.4}$$

where the first and last flow rates are set to $q_{j,0} = v_{j-1}$ and $q_{j,D_j} = v_j$. Each sub-segment *d* has a time interval of influence $\mathcal{T}^{\text{in}}_{F,j,d}$, where for the first and last segments it holds that $\mathcal{T}^{\text{in}}_{F,j,0} = [t_j - L/\gamma_{II}(v_{j-1}), t_j - L/\eta_{II}(v_j, q_{j,1})]$ and $\mathcal{T}^{\text{in}}_{F,j,D_j} = [t_j - L/\eta_{II}(q_{j,D_j-1}, v_j), t_j - L/\gamma_{II}(v_j)]$, while for intermediate segments $\mathcal{T}^{\text{in}}_{F,j,d} = [t_j - L/\eta_{II}(q_{j,d-1}, q_{j,d}), t_j - L/\eta_{II}(q_{j,d}, q_{j,d+1})]$, where $\eta_{II}(\cdot)$ is the backward shockwave speed. As a result, we need to consider all outflow rates in set $\mathcal{V}^{\text{in}}_F(t) = \left\{ v_{j-1}, q_{j,1}, \ldots, q_{j,D_j-1}, v_j \,|\, j \in \mathcal{J}_F(t) \right\}$ during fanning.

If we would have used D_j - step outer linearization to approximate $\overleftarrow{F}_j(t)$ we get for each $j \in \mathcal{J}_F(t)$ the following $D_j - 1$ projected linear fanning line segments, with $d = 1, \ldots, D_j - 1$,

$$\overleftarrow{F}^{\text{out}}_{j,d}(t \mid q_{j,1}, \ldots, q_{j,D_j}) = V(t_j) + \left(t - t_j + \frac{L}{\gamma_{II}\left(\phi_{II}(q_{j,d}, q_{j,d+1})\right)} \right) \phi_{II}(q_{j,d}, q_{j,d+1}) + \xi_{II}\left(\phi_{II}(q_{j,d}, q_{j,d+1})\right), \quad \text{if } t \in \mathcal{T}_{F,j,d}, \tag{C.5}$$

where the first and last flow rates are set to $q_{j,1} = v_{j-1}$ and $q_{j,D_j} = v_j$, intermediate flow rates $\phi_{II}(\cdot)$ are defined similar to (30), and $\mathcal{T}^{\text{out}}_{F,j,d} = [t_j - L/\gamma_{II}(q_{j,d}), t_j - L/\gamma_{II}(q_{j,d+1})]$. This simplifies the set of potential outflow rates that needs to be checked to $\mathcal{V}^{\text{out}}_F(t) = \left\{ \phi_{II}(v_{j-1}, q_{j,2}), \phi_{II}(q_{j,2}, q_{j,3}), \ldots, \phi_{II}(q_{j,D_j-1}, v_j) \mid j \in \mathcal{J}_F(t) \right\}$.

The potential inflow rate $\overleftarrow{v}(t)$ as defined in Equation (24) can be obtained by determining the flow rate that belongs to the line segment that minimises Equation (42).

Appendix D: Tracking shockwaves to determine potential flow rate changes

We aim to find the solution $\vec{t}_{i_1,i_2}$ that solves $\vec{U}_{i_1}(\vec{t}_{i_1,i_2}) = \vec{U}_{i_2}(\vec{t}_{i_1,i_2})$ with $i_1 < i_2$, where must hold that $\vec{t}_{i_1,i_2} \in \mathcal{T}_{i_1}$ and $\vec{t}_{i_1,i_2} \in \mathcal{T}_{i_2}$. Using the definition in Equation (33) it holds that time periods $\mathcal{T}_{i_1}$ and $\mathcal{T}_{i_2}$ only overlap if $u_{i_1} < u_{i_2}$. In this case, it holds that $\vec{t}_{i_1,i_2} \in [t_{i_2} + L/\gamma_I(u_{i_2}), t_{i_1+1} + L/\gamma_I(u_{i_1})]$, and $\vec{t}_{i_1,i_2}$ can be found by using Equation (38),

$$U(t_{i_1}) + \left(\vec{t}_{i_1,i_2} - t_{i_1} - \frac{L}{\gamma_I(u_{i_1})} \right) u_{i_1} + \xi_I(u_{i_1}) = U(t_{i_2}) + \left(\vec{t}_{i_1,i_2} - t_{i_2} - \frac{L}{\gamma_I(u_{i_2})} \right) u_{i_2} + \xi_I(u_{i_2}). \tag{D.1}$$

Substituting the definition of $\xi_I(q)$ stated in Equation (21) yields

$$\begin{aligned} U(t_{i_1}) + \left(\vec{t}_{i_1,i_2} - t_{i_1} - \frac{L}{\gamma_I(u_{i_1})} \right) u_{i_1} + L u_{i_1} \left(\frac{L}{\gamma_I(u_{i_1})} - \frac{1}{\sigma_I(u_{i_1})} \right) = \\ U(t_{i_2}) + \left(\vec{t}_{i_1,i_2} - t_{i_2} - \frac{L}{\gamma_I(u_{i_2})} \right) u_{i_2} + L u_{i_2} \left(\frac{L}{\gamma_I(u_{i_2})} - \frac{1}{\sigma_I(u_{i_2})} \right). \end{aligned} \tag{D.2}$$

Several terms cancel out, resulting in

$$U(t_{i_1}) + \left(\vec{t}_{i_1,i_2} - t_{i_1} \right) u_{i_1} - \frac{L u_{i_1}}{\sigma_I(u_{i_1})} = U(t_{i_2}) + \left(\vec{t}_{i_1,i_2} - t_{i_2} \right) u_{i_2} - \frac{L u_{i_2}}{\sigma_I(u_{i_2})}. \tag{D.3}$$

Since it holds that $\sigma_I(q) = q / \Phi_I^{-1}(q)$, we can rewrite Equation (D.3) as

$$U(t_{i_1}) + \left(\vec{t}_{i_1,i_2} - t_{i_1} \right) u_{i_1} - L\Phi_I^{-1}(u_{i_1}) = U(t_{i_2}) + \left(\vec{t}_{i_1,i_2} - t_{i_2} \right) u_{i_2} - L\Phi_I^{-1}(u_{i_2}). \tag{D.4}$$

Since it holds that

$$\begin{aligned} U(t_{i_2}) &= U(t_{i_2-1}) + (t_{i_2} - t_{i_2-1}) u_{i_2-1} \\ &= U(t_{i_2-2}) + (t_{i_2-1} - t_{i_2-2}) u_{i_2-2} + (t_{i_2} - t_{i_2-1}) u_{i_2-1} \\ &\vdots \\ &= U(t_{i_1}) + \sum_{i=i_1}^{i_2-1} (t_{i+1} - t_i) u_i \\ &= U(t_{i_1}) + t_{i_2} u_{i_2-1} - t_{i_1} u_{i_1} - \sum_{i=i_1+1}^{i_2-1} t_i (u_i - u_{i-1}), \end{aligned} \tag{D.5}$$

we can write Equation (D.4) as

$$\left(\vec{t}_{i_1,i_2} - t_{i_1}\right)u_{i_1} - L\Phi_I^{-1}(u_{i_1}) = t_{i_2}u_{i_2-1} - t_{i_1}u_{i_1} - \sum_{i=i_1+1}^{i_2-1} t_i(u_i - u_{i-1}) + \left(\vec{t}_{i_1,i_2} - t_{i_2}\right)u_{i_2} - L\Phi_I^{-1}(u_{i_2}). \tag{D.6}$$

Re-arranging terms results in

$$\vec{t}_{i_1,i_2} = \frac{\sum_{i=i_1+1}^{i_2} t_i(u_i - u_{i-1})}{u_{i_2} - u_{i_1}} + \frac{L}{\eta_I(u_{i_1}, u_{i_2})}. \tag{D.7}$$