

Graphical Solution and Continuum Approximation for the Single Destination Dynamic User Equilibrium Problem

Jorge A. Laval^{a,*}

^a*School of Civil and Environmental Engineering, Georgia Institute of Technology*

Abstract

This paper studies the user optimum dynamic traffic assignment in a network consisting of a congested freeway section connected to a surface street via capacitated off-ramps. It is shown that although the network is simple, it is in general complicated to obtain the optimum solution analytically because (i) it has two different alternating diversion patterns, and (ii) it can only be obtained sequentially in time. A simplified graphical solution method is proposed to cope with (i) but the method remains sequential. Finally, a continuum approximation is proposed that enables finding closed-form solutions without the above limitations.

Key words:

User optimum, dynamic traffic assignment, continuum approximation

1 Introduction

Currently, most dynamic user optimum (DUO) traffic assignment models are either based on link performance functions (Bliemer and Bovy, 2003; Han, 2003; Bellei et al., 2005; Jang et al., 2005) or traffic simulation (Lo and Szeto, 2002; Peeta and Zhou, 2006; Jin, 2007) to compute link travel times at each time instant. While the drawbacks of using link performance functions are now well understood (Daganzo, 1995; Nie and Zhang, 2005), the simulation approach still relies on numerical methods for solving the traffic dynamics and the optimization problem, which prevents us from gaining meaningful insights.

* Corresponding author. Tel. : +1 (404) 894-2360; fax:+1 (404) 894-2278
Email address: jorge.laval@ce.gatech.edu (Jorge A. Laval).

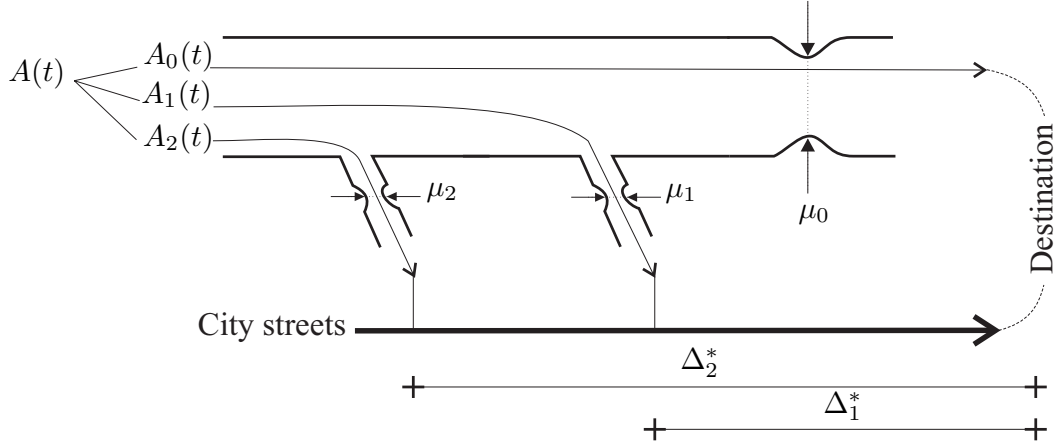


Fig. 1. Schematic representation of the network.

Although unable to cope with queue spillovers, a third approach, the so-called “point-queue models”, provides a realistic representation of traffic delays and allows for analytical solutions in simple networks. Analytical solutions have been found for the social optimum dynamic traffic assignment (Muñoz and Laval, 2005; Shen et al., 2007) and for the DUO with departure time choice (see Lago and Daganzo, 2007, and references therein) in the case of simplified networks. To the author’s knowledge no analytical formulae has been derived for the DUO case.

In this paper, we use the point-queue approach to solve analytically the single destination DUO problem. To this end, section 2 describes the modeling assumptions and the type of networks considered here, i.e., multiple capacitated parallel routes. Section 3 shows that the user optimal solution consists of two different assignment patterns that alternate in time, and this makes it difficult to find closed-form solutions. This section also introduces a graphical solution method that is effective when demand increases rapidly at the beginning of the rush. Closed-form solutions are obtained using a continuum approximation in section 4. Finally, a discussion is presented in section 5.

2 Problem formulation

Suppose that we know the cumulative count curve $A(t)$ of vehicles entering a freeway segment with R off-ramps and a bottleneck of capacity μ_0 ; see Fig. 1 for the case $R = 2$. Vehicles are bound for a single destination past the bottleneck, which may be bypassed by taking any of the $r = 1 \dots R$ off-ramps heading to the local streets grid. This diversion is at the expense of extra free-flow travel time and possibly some queuing due to off-ramp r ’s capacity, μ_r , which may be dictated by its discharge capacity to the city streets.

Without loss of generality, it is assumed that (i) the freeway free-flow travel time is zero since the relevant characteristic of a route is its *extra* travel time with respect to the freeway's free-flow travel time; and (ii) that the arrival pattern is unimodal, where the arrival rate continuously increases to exceed freeway capacity, and then decreases to values under capacity.

This problem defines $R + 1$ alternative routes, with route $r = 0$ being the freeway and route $r = 1 \dots R$ representing the route of a vehicle taking off-ramp r ; see Fig. 1. As shown in the figure, the cumulative count curve of vehicles using route r is denoted $A_r(t)$. In the sequel we will use a dot on top of a variable to denote time differentiation; e.g., $\dot{A}_r(t)$ is the flow on route r at time t .

Let $\Delta_r(t)$ be the trip time in route r at time t :

$$\Delta_r(t) = \Delta_r^* + w_r(t),$$

where Δ_r^* is a constant travel time-independent of flow-incurred when taking off-ramp r and $w_r(t)$ is the corresponding queuing delay. Naturally, $\Delta_r^* < \Delta_{r+1}^*$ and $\Delta_0^* = 0$. It is straightforward to show that if route r is congested then the delay is given by

$$w_r(t) = \frac{A_r(t) - A_r(t_r)}{\mu_r} - (t - t_r), \quad r = 0 \dots R(t), \quad (1)$$

where t_r represents the time when route r begins to be congested (i.e., when the arrival rate exceeds capacity), and $R(t)$ is the most upstream off-ramp being congested by time t . Of course, $w_r(t) = 0$ if $r > R(t)$.

According to the DUO equilibrium condition, the travel time in equilibrium, $\Delta(t)$, satisfies

$$\text{UO at time } t \Rightarrow \begin{cases} \Delta(t) = \Delta_0(t) = \Delta_1(t) = \dots = \Delta_{R(t)}(t) \leq \Delta_{R(t)+1}(t) \\ \Delta(t) < \Delta_{R(t)+2}(t), \Delta_{R(t)+3}(t) \dots \end{cases} \quad (2)$$

The goal is to determine the time dependent paths vehicles should follow so that (2) is satisfied; i.e., so that individual travel times are minimized at each time instant.

3 Analytical and graphical solutions

To simplify the exposition we start with the single off-ramp case shown in Fig. 2, which depicts the queuing diagrams for both the freeway and the off-

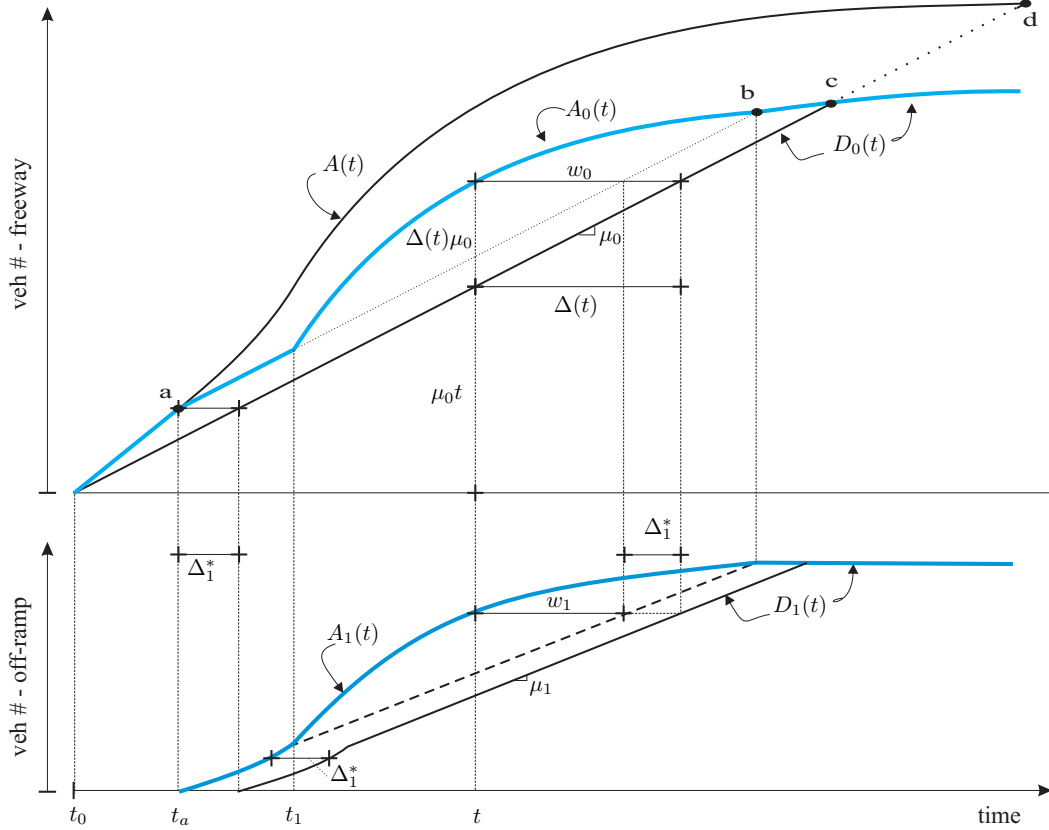


Fig. 2. Queuing diagrams for the freeway and one off-ramp satisfying the UO equilibrium.

ramp under UO equilibrium. The cumulative departure curve of vehicles using route r , denoted $D_r(t)$, can be interpreted as the cumulative arrivals to the destination using each route. The time axis starts at $t = t_0$ when congestion starts in the freeway; i.e., the earliest time when $\dot{A}(t) = \mu_0$.

Under UO all incoming traffic will stay in the freeway until point “a” in the figure, when the delay $w_0(t)$ equals the free-flow travel time in route 1, Δ_1^* . At this point “a” incoming flow in excess of the freeway capacity μ_0 should divert to the off-ramp in order to maintain a constant freeway delay of Δ_1^* ; e.g., as in the interval $[t_a, t_1]$ in the figure. This “type-1” diversion pattern can be sustained up to t_1 when the excess flow exceeds the off-ramp capacity and a queue starts to form. Therefore, starting at $t = t_1$ a different “type-2” diversion pattern will be observed, where a certain proportion of incoming traffic will divert to route 1 such that (2) is satisfied; i.e., such that $w_0(t) = w_1(t) + \Delta_1^*$ as exemplified in the figure. This diversion pattern will continue up to point “b” in the figure where the queue dissipates in the off-ramp. At this point all incoming traffic should use the freeway until the end of the rush hour at point “c” where the queue in the freeway dissipates.

To summarize, and generalizing for R off-ramps, type-1 diversion patterns will

be observed whenever (i) the equilibrium travel time equals the free-flow travel time in the most downstream off-ramp; i.e., $\Delta(t) = \Delta_{R(t)+1}^*$ and (ii) the excess demand $\dot{A}(t) - \sum_{k=0}^{R(t)} \mu_k$ is less than the capacity of off-ramp $R(t) + 1$. In this case the arrival flow to all off-ramps is μ_r except for $R(t) + 1$ which receives the excess demand; i.e.,

$$\dot{A}_r(t) = \begin{cases} \mu_r & \text{if } r \leq R(t), \\ \dot{A}(t) - \sum_{k=0}^{R(t)} \mu_k & \text{if } r = R(t) + 1, \\ 0 & \text{o/w.} \end{cases} \quad (\text{type-1 diversion}) \quad (3)$$

The arrival flows for type-2 diversion patterns will be observed whenever $\Delta(t) < \Delta_{R(t)+1}^*$. Solving the system of equations (2)-(1) quickly becomes complicated due to all the history of the system that must be carried over to a given time instant. A much more elegant derivation follows from using the DUO condition in differential form; i.e.,

$$\dot{\Delta}_0(t) = \dot{\Delta}_1(t) = \dots = \dot{\Delta}_{R(t)}(t). \quad (4)$$

Note that this differential form is always true but useful only in type-2 diversion, where all the off-ramps being used receive an inflow greater than their capacity; in type-1 diversion all travel time rates are zero. The system of equations (1) and (4) leads to the following solution:

$$\dot{A}_r(t) = \begin{cases} \frac{\mu_r}{\sum_{k=0}^{R(t)} \mu_k} \dot{A}(t) & \text{if } r \leq R(t), \\ 0 & \text{o/w.} \end{cases} \quad (\text{type-2 diversion}) \quad (5)$$

Interestingly, it can be seen that the flow using each route is a constant fraction of the total arrival flow, and that this fraction is proportional to the capacity of the route.

The nature of the solution presented in this section implies that it is in general complicated to obtain the optimum solution analytically. This is because (i) there are two different diversion patterns that alternate in time as new off-ramps are being used; and (ii) the solution can only be obtained sequentially in time since one needs to know $R(t)$ and the prevailing diverging pattern to estimate the flows on each route. In this sequential solution method, one can proceed as outlined in Algorithm 1; see the appendix. Although this algorithm should be effective, a streamlined graphical solution method can be devised whenever type-one diversion patterns can be neglected, as shown next.

3.1 Simplified graphical solution method

In this section we propose a simple graphical solution method that can be applied when the arrival curve increases rapidly at the beginning of the rush hour. Specifically, we require that the excess demand $\dot{A}(t) - \sum_{k=0}^{R(t)} \mu_k$ be always greater than the capacity of the most upstream off-ramp being utilized, $\mu_{R(t)}$. In this situation the graphical solution is greatly simplified (compared to the “standard” approach used in Fig. 2) because one can neglect type-1 diversion patterns. This allows us to integrate (5) and obtain a tractable solution; i.e.:

$$\begin{aligned}
A_r(t) &= \int_0^t \Xi_{R(s)}^r \dot{A}(s) ds \\
&= \sum_{p=r}^{R(t)-1} \int_{t_p}^{t_{p+1}} \Xi_p^r \dot{A}(s) ds + \int_{t_{R(t)}}^t \Xi_{R(s)}^r \dot{A}(s) ds \\
&= \sum_{p=r}^{R(t)-1} \Xi_p^r [A(t_{p+1}) - A(t_p)] + \Xi_{R(t)}^r [A(t) - A(t_{R(t)})], \\
&= \mu_r (t_{R(t)} - t_r + \Delta_{R(t)}^* - \Delta_r^*) + \Xi_{R(t)}^r [A(t) - A(t_{R(t)})], \tag{6}
\end{aligned}$$

where

$$\Xi_i^r = \frac{\mu_r}{\sum_{k=0}^i \mu_k}$$

is the fraction of the total arrivals diverted to route r when there are i ramps being utilized. Equality (6) follows after noting that

$$A(t_r) - A(t_{r-1}) = [(t_r - t_{r-1}) + (\Delta_r^* - \Delta_{r-1}^*)] \sum_{k=0}^{r-1} \mu_k. \tag{7}$$

One can verify that (7) is true by considering Fig. 2, where it can be seen that $A_0(t_1) - A_0(t_0) = [(t_1 - t_0) + (\Delta_1^* - \Delta_0^*)] \mu_0$. In general, this can be written as $A_k(t_r) - A_k(t_{r-1}) = [(t_r - t_{r-1}) + (\Delta_r^* - \Delta_{r-1}^*)] \mu_k$, for all route $k = 0 \dots r$. Since at time t_r there are $r - 1$ routes carrying flow, then $A(t_r) = \sum_{k=0}^{r-1} A_k(t_r)$ and (7) follows.

Finally, the main result in this section is obtained by replacing (6) in (1); namely,

$$w_r(t) = \frac{A(t) - A(t_{R(t)})}{\sum_{k=0}^{R(t)} \mu_k} - (t - t_{R(t)}) + (\Delta_{R(t)}^* - \Delta_r^*), \quad r = 0 \dots R(t). \tag{8}$$

We now show that (8) implies that queuing delays can be presented graphically

using total arrivals $A(t)$ and total departures, $D(t)$, defined as

$$D(t) = \sum_{k=0}^{R(t)} \mu_k.$$

To see this note that under UO delays can also be written as

$$w_r(t) = \frac{A_r(t) - A_r(t_{R(t)})}{\mu_r} - (t - t_{R(t)}) + (\Delta_{R(t)}^* - \Delta_r^*), \quad r = 0 \dots R(t). \quad (9)$$

The reader can verify that this is true with the help of Fig. 2, where it can be seen that $t_0 = t_1 - (A_0(t_1)/\mu_0 - \Delta_1^*)$. This relationship can be generalized to $t_r = t_R - (A_r(t_R)/\mu_r - \Delta_R^*) + \Delta_r^*$. Replacing this last equality in (1) gives (9) as sought. Comparing (8) with (9) indicates that indeed system delays can be represented graphically using total arrivals and total departures.

The resulting simplified graphical solution method is presented in Fig. 3. We start by drawing the total arrivals $A(t)$ for $t \geq t_0$ and a straight line of slope μ_0 from the origin, which corresponds to the first segment of $D(t)$. Route 1 will start being used as soon as the horizontal distance between arrivals and departures equals Δ_1^* ; this defines t_1 . This implies that the total departure rate increases in μ_1 , and therefore the slope of the departure curve changes to $\mu_0 + \mu_1$ at time $t_1 + \Delta_1^*$; see Fig. 3. This process is repeated until the arrival flow drops below the departure rate, at which point the queues in each route will begin to dissipate, upstream routes first. The time route r 's queue dissipates is given by the time when the horizontal distance between arrivals and departures reaches Δ_r^* for the second time. At this point, no flow will diverge to route r and therefore the slope of the departure curve must be decreased by μ_r . This process continues until all queues have vanished. This graphical solution method can be streamlined using the Algorithm 2 in the appendix.

Notice that all relevant information can be extracted from Fig. 3. For example, the travel time in each route is the horizontal distance between arrivals and departures, as usual, which is the same for all routes. Therefore, the area under the curve is the total travel time in the system and the vertical distance between the curves is the number of vehicles being delayed in the system. The queuing delay for route r can be obtained by shifting the departure curve horizontally to the left by Δ_r^* ; see thin lines in Fig. 3. As shown in the figure, the queuing delay $w_r(t)$ is simply the horizontal distance between the arrival curve and the corresponding shifted departure curve. The figure also shows the total costs associated with each route being used; light gray for the freeway, gray for off-ramp 1 and dark gray for off-ramp 2. It becomes evident from the figure that a higher proportion of total system cost is spent in the freeway, and that this proportion decreases with r .

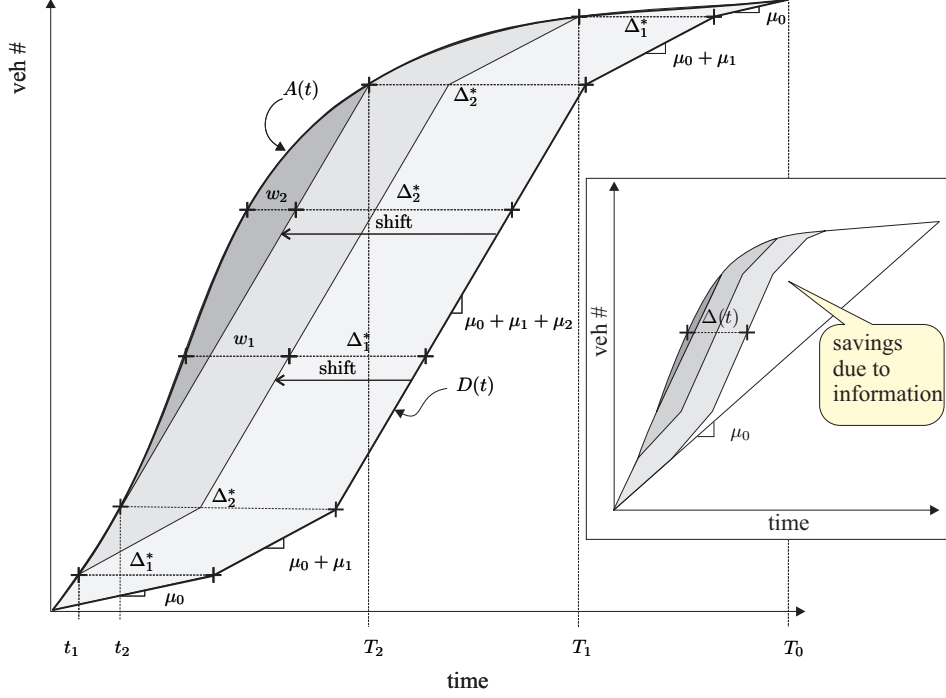


Fig. 3. Simplified graphical solution method when 2 off-ramps are used.

Although the graphical solution method presented in this section is simple to apply, it does not provide tractable analytical solutions. Fortunately, the continuum approximation method presented next overcomes this limitation.

4 Continuum approximation

In this section we use a continuum approximation (CA) of the set of discrete off-ramps where vehicles can exit the freeway at any given location $x \geq 0$ upstream of the freeway bottleneck. The freeway bottleneck is located at $x = 0$ and has a capacity μ_0 . This approximation is characterized by an “information wave” that marks the most upstream location where vehicles divert. Let $x = \xi(t)$ be the trajectory of this information wave. Since vehicles are diverting in $0 \leq x \leq \xi(t)$, the total lateral diverting flow at time t , $L(t)$, can be expressed as:

$$L(t) = \phi \xi(t), \quad (10)$$

where ϕ is a constant representing the lateral exit capacity in units of vehicles / (time \times distance). Let $A_L(t)$ be the cumulative lateral flow by time t , namely:

$$A_L(t) = \int_0^t L(s) ds. \quad (11)$$

Total arrivals can be decomposed in vehicles staying in the freeway and vehicles diverging laterally; i.e., $A(t) = A_0(t) + A_L(t)$. Here, $t' \leq t$ is a non-trivial function of t , which turns the problem mathematically intractable. However, if the total arrival flow varies slowly over time one can postulate:

$$A(t) = A_0(t) + A_L(t), \quad (12)$$

which will allow us to identify the ordinary differential equation (ODE) defining the dynamics of the system. The reader can verify (e.g., by looking at the top diagram in Fig. 2a along the vertical line at time t) that freeway cumulative arrivals can be expressed as $A_0(t) = \mu_0 \Delta(t) + \mu_0 t$. To express $A_L(t)$ in terms of $\Delta(t)$ it is assumed that:

$$\Delta(t) = \xi(t)/v_A, \quad (13)$$

where v_A should be comparable to the average speed in local streets. Eliminating $\xi(t)$ from (10) and (13) and in combination with (11) gives $A_L(t) = \phi v_A \int_0^t \Delta(s) ds$. It follows that system dynamics are given by $\Delta(t)\mu_0 + \phi v_A \int_0^t \Delta(s) ds - (A(t) - \mu_0 t) = 0$, or equivalently by:

$$\Delta(t) + a \int_0^t \Delta(s) ds - d(t) = 0, \quad (14)$$

where $a = \phi v_A / \mu_0$ is a constant with units of time^{-1} and $d(t) = A(t) / \mu_0 - t$ can be interpreted as the queuing delay in the absence of diversion. Eqn. (14) is a first-order ODE in $y(t) = \int_0^t \Delta(s) ds$, whose general solution satisfies

$$y(t) = \int_0^t d(s) e^{-a(t-s)} ds. \quad (15)$$

It is interesting to note that the behavior of the system described in this section is governed by a single parameter, a . Typical values for a should be in the range $[5 - 15]$, considering that $\phi \approx 1,000$ veh/hr/km, $v_A \approx 40$ km/hr, and $\mu_0 \in [4,000 - 10,000]$ veh/hr. Naturally, the precise analytical solution of (14) depends on the functional form of the total arrival curve, or more precisely, on the ratio $A(t)/\mu_0$.

4.1 Examples

Here we show three families of arrival curves for which the CA produces simple yet insightful analytical solutions. We also compare these results with the “exact” discrete solution using Algorithm 2 in the appendix. For this comparison it is assumed that $\mu_0 = 4,000$ veh/hr, $v_A = 30$ km/hr and that there is one off-ramp per km with capacity $\mu_r = 1,000$ veh/hr, $r = 1, 2, \dots$. The latter implies $a = 7.5$ hr⁻¹ for the CA solution, which is the value used throughout the following examples unless otherwise indicated.

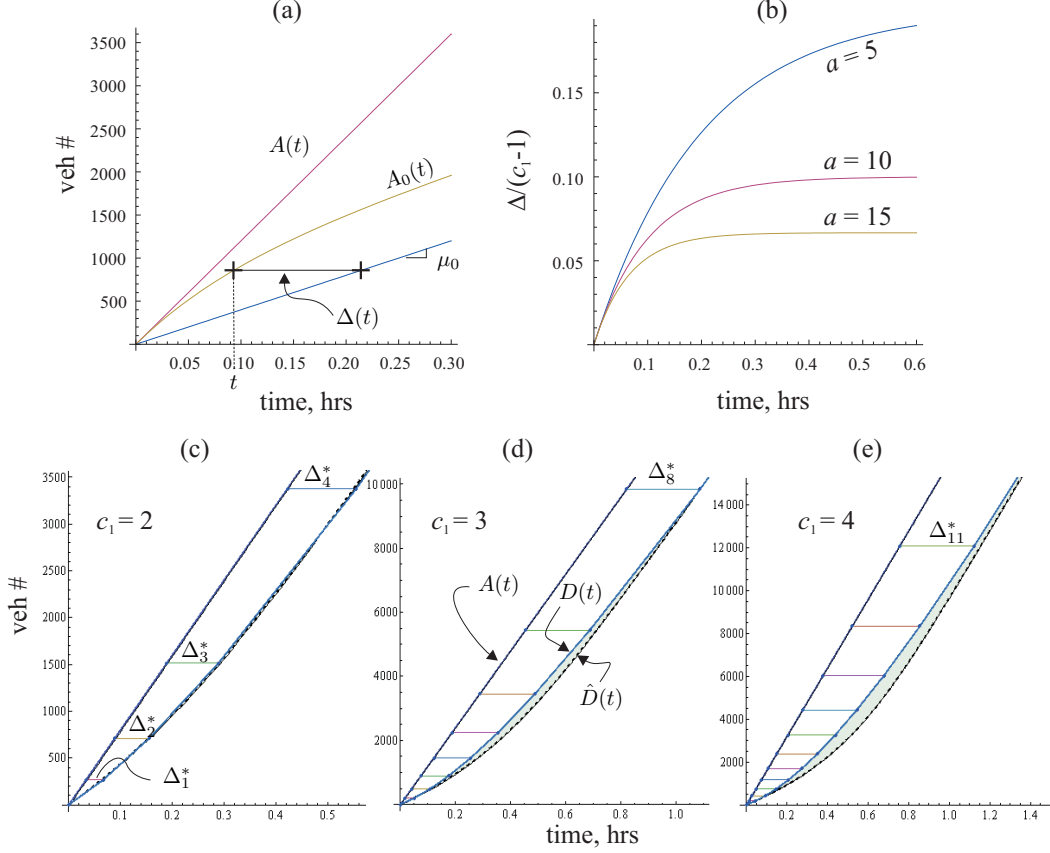


Fig. 4. CA solution with linear arrivals: (a) freeway queuing diagram ($c_1 = 3$); (b) normalized travel time for several values of a ; (c–e) comparison with exact solution in terms of total departure curves.

4.1.1 Linear arrivals

Although a linear arrival curves satisfying $A(t)/\mu_0 = c_1 t$, $c_1 > 1$ is certainly not realistic to represent the entire rush hour period, it does provide useful insights on system behavior by the beginning of the rush. In this case the delay without diversion is given by $d(t) = (c_1 - 1)t$ and the solution of (14) becomes

$$\Delta(t) = \frac{c_1 - 1}{a}(1 - e^{-at}), \quad (16)$$

which stabilizes at $(c_1 - 1)/a$ as $t \rightarrow \infty$. According to (13) this implies that the information wave stabilizes at a well-defined location upstream of the freeway bottleneck; i.e., $\xi(t \rightarrow \infty) = v_A(c_1 - 1)/a$.

Figure 4a shows the solution in terms of the freeway's queuing diagram, where it can be seen that $A_0(t)$ increases rapidly at the beginning of the rush and gradually decay towards the bottleneck discharge rate μ_0 . Similarly, in the discrete case off-ramp flows (5) can be approximated by $\mu_r/(\mu_0 + L(t))c_1\mu_0 = c_1\mu_r/(1 + a\Delta(t))$, which also tends to the bottleneck discharge rate μ_r . Part

b of the figure depicts $\Delta(t)/(c_1 - 1)$ for several values of a , which shows that all relevant variables stabilize to equilibrium values shortly after the start of the rush.

Figures 4c-e compare the CA solution and the exact discrete solution. The total departure curve using the CA method is labeled $\hat{D}(t)$ in the figure, while $D(t)$ represents the exact solution. It can be seen how the approximation is quite accurate for low values of c_1 but deteriorates as c_1 grows. Notice, however, that in steady-state both solutions are identical; i.e, in both cases $\Delta(\infty) = (c_1 - 1)/a$.

4.1.2 Exponential decay

A slightly more realistic arrival curve is such that the arrival rate decreases gradually over time. This can be accomplished in several ways, but simple analytical solutions can be obtained assuming that the arrival rate decays exponentially from $\mu_0(1 + \rho ad_0)$ at $t = 0$ to μ_0 , such that $d(t) = d_0(1 - e^{-\rho at})$; $\rho, d_0 > 0$. The parameter d_0 can be interpreted as the maximum delay in the absence of diversion, and ρ is a dimensionless parameter that controls the rate of decay. In this case the solution of (14) becomes

$$\Delta(t) = \begin{cases} d_0 at e^{-at} & \text{if } \rho = 1, \\ d_0 \rho (e^{-at} - e^{-\rho at}) / (\rho - 1) & \text{if } \rho \neq 1, \end{cases}$$

which tends to zero as $t \rightarrow \infty$, and tends to $d_0 e^{-at}$ as $\rho \rightarrow \infty$. This indicates that a sudden surge in the arrival flow (large ρ) translates into a sudden increase in travel time that decays to zero exponentially in time. The queuing diagram in Fig. 5a, exemplifies how this type of arrival curve produces freeway flows \dot{A}_0 that increase rapidly at the beginning of the rush followed by a gradual decay towards the bottleneck discharge rate. Figure 5b shows $\Delta(t)/d_0$, where it can be seen that the equilibrium travel time—and therefore also $\xi(t)$ —exhibit a “bell”-type shape with a peak at time t^* shortly after the beginning of the analysis period. Solving $\dot{\Delta}(t^*) = 0$ yields:

$$t^* = \begin{cases} 1/a & \text{if } \rho = 1, \\ 1/a \ln(\rho) / (\rho - 1) & \text{if } \rho \neq 1, \end{cases} \quad (17)$$

which does not depend on the parameter d_0 , but only on a and ρ . The maximum travel time during the rush hour is obtained by evaluating (17) at $t = t^*$, namely:

$$\Delta(t^*) = \begin{cases} d_0 e^{-1} & \text{if } \rho = 1, \\ d_0 \rho^{-1/(\rho-1)} & \text{if } \rho \neq 1. \end{cases} \quad (18)$$

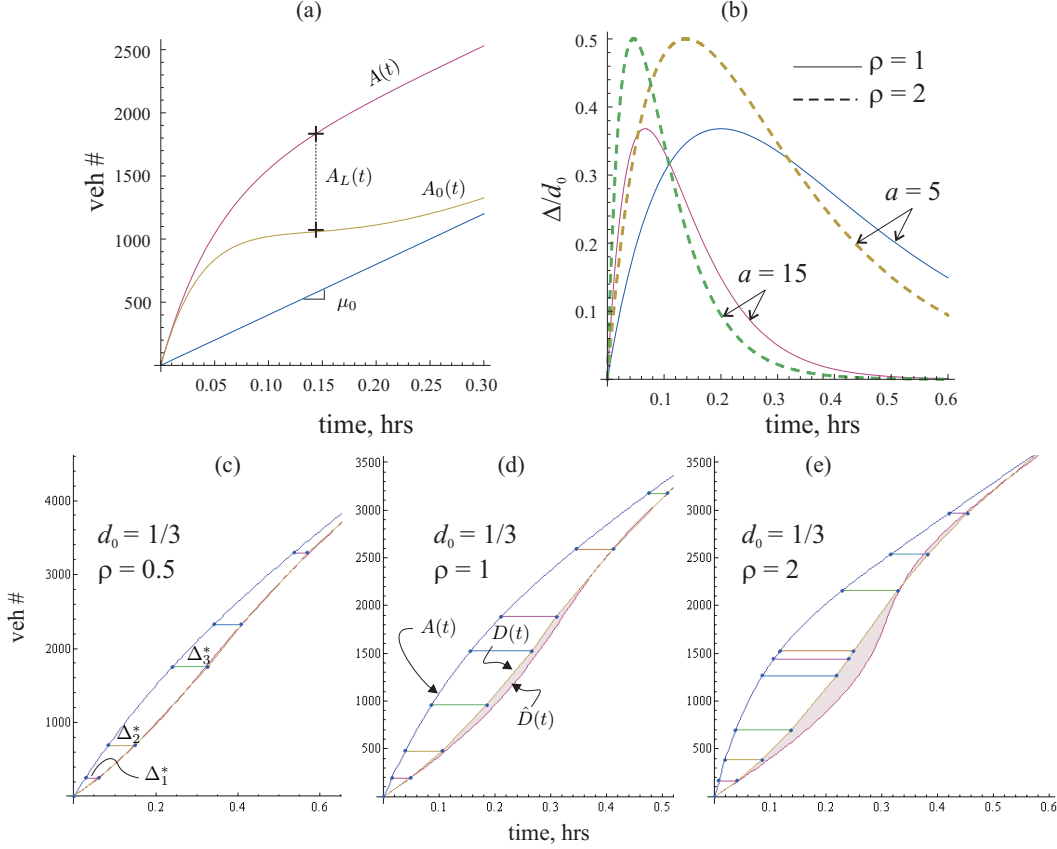


Fig. 5. CA solution with exponentially decaying arrivals: (a) freeway queuing diagram ($d_0 = 1/3$ hr and $\rho = 2$); (b) normalized travel time for several values of a and ρ ; (c–e) comparison with exact solution in terms of total departure curves.

Note that the maximum travel time relative to the maximum travel time without diversion $\Delta(t^*)/d_0$ is only a function of ρ . A typical value of ρ should be between 0.5 and 2, yielding maximum travel times that are between a fourth and a half of the maximum travel time without diversion, respectively.¹

Similarly to the linear case, Figs. 5c-e show that the method is more accurate for lower arrival rates, and that in steady-state the approximation is the exact.

4.1.3 Piecewise linear arrivals

Piecewise linear arrivals are useful because any arrival pattern can be approximated arbitrarily well by increasing the number of linear segments. In this

¹ Typical values for ρ are obtained by assuming that the time interval where the arrival flow exceeds the bottleneck capacity by at least 1% is approximately 1 hour, for values of a between 5 and 15.

case the total arrival flow is constant by segments; i.e.,

$$\dot{A}(t) = c_i \mu_0, \quad T_{i-1} < t \leq T_i, i = 1, 2, \dots,$$

with $T_0 = 0$ and $c_0 = 1$. Replacing $A(t) = \int_0^t \dot{A}(s) ds$ in (14) gives:

$$\Delta(t) = \frac{1}{a} (c_i - 1 + \sum_{j=1}^i (c_{j-1} - c_j) e^{-a(t-T_{j-1})}), \quad T_{i-1} < t \leq T_i, i = 1 \dots i^*, \quad (19)$$

where i^* is such that the queue vanishes at T_{i^*} ; and $\Delta(t) = 0$ if $t \geq T_{i^*}$. The reader can verify that for the first segment $0 \leq t \leq T_1$ eqn. (19) simplifies to (16), as expected. The CA for piecewise linear arrivals is exemplified in the next section for the bilinear case ($i=2$), which also shows the conditions for stability.

4.2 Stability

The CA method is accurate so long as arrival flows vary slowly over time. In our case we find that $\dot{A}(t)$ should decrease slowly at the end of the rush, at least slowly enough to preclude the appearance of spurious results. Spurious results are a consequence of violating the condition

$$\dot{\Delta}(t) \geq -1, \quad (20)$$

which states that the variation in delay cannot exceed the variation in time itself. To see the effects of violating this condition, Fig. 6 shows a bilinear arrival curve $A(t)$ with $c_2 \ll c_1$, and the corresponding exact solution, $D(t)$, obtained with Algorithm 2. It can be seen that the total departure curve predicted by (19), labeled $\hat{D}(t)$ in the figure, degenerates shortly after T_1 .

Unfortunately, it is not possible to find a general closed-form condition for $\dot{A}(t)$ to satisfy (20). To see this, one can take derivatives of (14) and combine it with (20) to obtain $\dot{A}(t)/\mu_0 \geq a\Delta(t) - 1$, which depends on $\Delta(t)$. However, for the arrival curves presented here it is possible to find the stability condition by identifying the time t^* such that $\dot{\Delta}(t^*)$ is minimum, and solving (20) at t^* . For exponential decaying arrivals this exercise shows that the stability condition is $ad_0 \leq r^{\frac{2}{r-1}}$ if $r \neq 1$ and $ad_0 \leq e^2$ if $r = 1$; for bilinear arrivals the condition is $c_2 \geq (c_1 - 1)(1 - e^{-aT_1})$.

From a practical standpoint, the stability condition for piecewise linear arrivals can be quite restrictive. To overcome this limitation, one can “round the edges” of $A(t)$ by smoothing out the transition between different flow levels. This can be done in a number of ways; the appendix shows the logistic smoothing technique, which still allows for analytical (but generally non-algebraic) solutions.

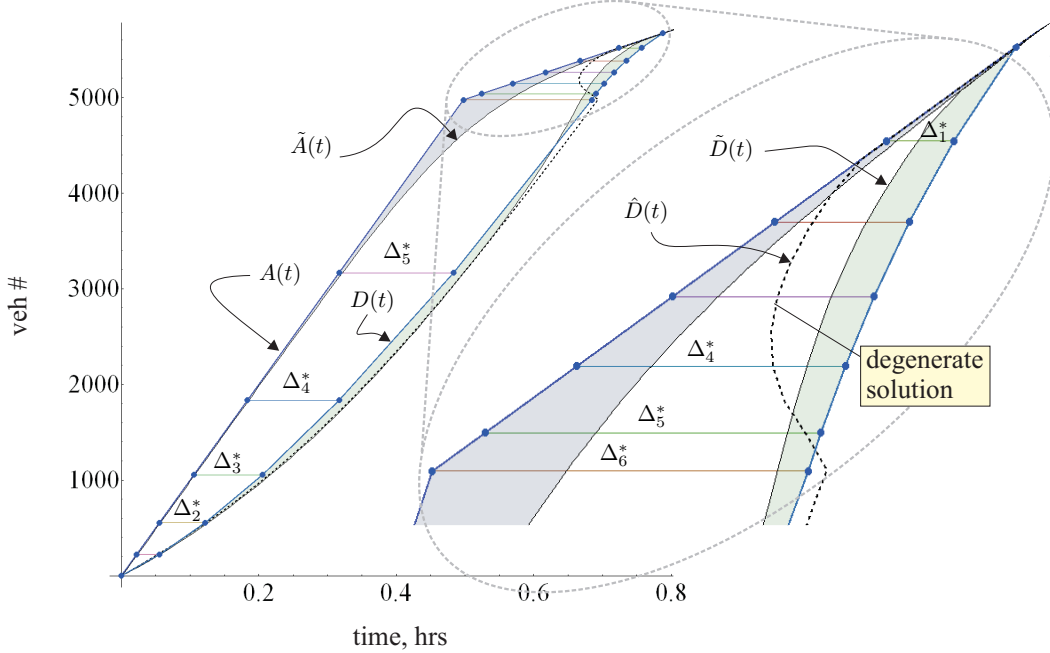


Fig. 6. Comparison between the exact solution $D(t)$, the CA solution $\hat{D}(t)$, and the solution using logistic smoothing, $\tilde{d}(t)$. The arrival curve parameters used are $c_1 = 5, c_2 = 0.5$ and $T_1 = 0.5$ hr, which induce a degenerate CA solution.

The total departure curve using logistic smoothing, $\tilde{D}(t)$, is also shown in Fig. 6, where it can be seen that the solution is physically well defined. Shaded areas in the figure are defined between $A(t)$ and the smooth arrival curve $\tilde{A}(t)$, as well as $D(t)$ and $\tilde{D}(t)$, in order to see the discrepancies due to logistic smoothing.

5 Discussion

The main assumption in this paper is that the network will not be subject to spillbacks, which may not be the case under heavily congested situations. Fortunately, the effects of this assumption are less deleterious than expected for the particular problem examined here. First, consider the spillback from the freeway into its exits. It turns out that the “information wave” $\xi(t)$ always travels faster than the back of the queue on the freeway. This is true because the trajectory of the back of the queue, $B(t)$, can be well approximated by the ratio between the number of vehicles in the queue, $A_0(t) - \mu_0 t$, and the density in the queue μ_0/v_F , where v_F is the speed in the freeway queue. From Fig. 2 one can see that $A_0(t) - \mu_0 t = \Delta(t)\mu_0$ and since $\xi(t) = \Delta(t)v_A$ we have that

$$\frac{B(t)}{\xi(t)} = \frac{v_F}{v_A},$$

which is always less than unity since $v_F < v_A$ (otherwise no one would divert). It follows that there will always be a proportion $1 - v_F/v_A$ in the freeway segment $0 \leq x \leq \xi(t)$, where spillbacks do not constrain vehicle diversion. Even in the segment subject to spillback, diversion would be constrained only when the desired exit rate at a particular off-ramp exceeds μ_0 , which should be extremely unlikely by virtue of (5).

Spillbacks from the off-ramps into the freeway are also less likely than expected, again by virtue of (5). As noted earlier, this diversion pattern assigns flows to routes proportionally to their capacity and inversely proportional to the number of routes being used. For example, in the worst case scenario of linear arrivals it can be shown that an upper bound for the number of vehicles in an off-ramp queue is $\ln(c_1)\mu_r/a$, which only grows logarithmically with c_1 . Naturally, there is no guarantee that spillbacks will not occur; an extended methodology that accounts for spillbacks is currently being investigated.

Although not considered here explicitly for clarity, on-ramps can be easily incorporated using the method proposed in (Muñoz and Laval, 2005), where the demand from all the origins—including on-ramps—is aggregated in the total arrival curve. This approach is valid so long as no on-ramp becomes a bottleneck. However, if this should be the case, one can divide the freeway corridor in segments with a single bottleneck and apply the methodology presented here to each segment.

The congestion on the city streets is only captured at the exit ramps, which have limited capacity. Unfortunately, even the simple extension of introducing a bottleneck common to all alternative routes turns the problem mathematically intractable due to the time lags involved. Note, however, that for practical and/or simulation applications where one can estimate the actual travel times Δ_r^* 's in real time, one can still apply the solution methods of §3 since they are sequential in time.

The framework proposed in this paper can be applied to evaluate the benefits of providing user information. Imagine that an incident takes place in an otherwise uncongested freeway and that drivers—unaware of the travel times on city streets—decide to stay in the freeway. In this case the total delay would be given by the area between the arrival and departure curves in the top of Fig. 2; i.e., the area from the origin up to point “d” in the figure. If user information is provided, cost savings should equal the white area shown in the inset of Fig. 3, as long as one or more off-ramps are used. As a rough approximation, cost savings should be around 50 to 75% as suggested by the result below eqn. (18).

The continuum approximation proposed here can be used to obtain the system optimum analytical solutions to the problem presented here. One can supple-

ment the graphical solution method presented in (Muñoz and Laval, 2005) with the corresponding expression for lateral outflows, $L(t)$. It turns out that in system optimum, $L(t)$ increases linearly in time, and therefore it is expected that the analytical solution should be simple to obtain. This is currently being investigated by the author.

Acknowledgment. The author is grateful to Juan Carlos Muños from Universidad Católica de Chile for his helpful comments and suggestions, and to two anonymous referees whose observations greatly improved the final form of this paper.

References

- Bellei, G., Gentile, G., Papola, N., 2005. A within-day dynamic traffic assignment model for urban road networks. *Transportation Research Part B: Methodological* 39 (1), 1–29.
- Bliemer, M., Bovy, P., 2003. Quasi-variational inequality formulation of the multiclass dynamic traffic assignment problem. *Transportation Research Part B: Methodological* 37 (6), 501–519.
- Daganzo, C., 1995. Properties of link travel time functions under dynamic loads. *Transportation Research Part B: Methodological* 29 (2), 95–98.
- Han, S., 2003. Dynamic traffic modelling and dynamic stochastic user equilibrium assignment for general road networks. *Transportation Research Part B: Methodological* 37 (3), 225–249.
- Jang, W., Ran, B., Choi, K., 2005. A discrete time dynamic flow model and a formulation and solution method for dynamic route choice. *Transportation Research Part B: Methodological* 39 (7), 593–620.
- Jin, W., 2007. A dynamical system model of the traffic assignment problem. *Transportation Research Part B: Methodological* 41 (1), 32–48.
- Lago, A., Daganzo, C., 2007. Spillovers, merging traffic and the morning commute. *Transportation Research Part B: Methodological* 41 (6), 670–683.
- Lo, H., Szeto, W., 2002. A cell-based variational inequality formulation of the dynamic user optimal assignment problem. *Transportation Research Part B: Methodological* 36 (5), 421–443.
- Muñoz, J., Laval, J. A., 2005. System optimum dynamic traffic assignment graphical solution method for a congested freeway and one destination. *Transportation Research Part B* 40 (1), 1–15.
- Nie, X., Zhang, H., 2005. Delay-function-based link models: their properties and computational issues. *Transportation Research Part B: Methodological* 39 (8), 729–751.
- Peeta, S., Zhou, C., 2006. Stochastic quasi-gradient algorithm for the off-line stochastic dynamic traffic assignment problem. *Transportation Research Part B: Methodological* 40 (3), 179–206.

Shen, W., Nie, Y., Zhang, H., 2007. On path marginal cost and its relation to dynamic system-optimal traffic assignment. In: Heydecker, B., Bell, M., Allsop, R. (Eds.), 17th International Symposium on Transportation and Traffic Theory. Elsevier, New York.

A Algorithms

This appendix provides algorithms that can help automate the graphical solution methods outlined in §3. As in the main text, t_0 below stands for the time when congestion begins in the freeway.

The sequential solution method to obtain the exact solution in §3 can be stated as follows.

Algorithm 1 *Set $t = t_0$ and $R(t_0) = 0$. Then perform the following steps:*

- Step 1.** *assign flows to each route $r \leq R(t)$ according to (6) until t' when $\Delta(t') = \Delta_{R(t)+1}^*$ or $\Delta(t') = \Delta_{R(t)-1}^*$,*
- Step 2.** *if $\Delta(t') = \Delta_{R(t)+1}^*$ then set $R(t') = R(t) + 1$, and assign flows to each route according to (4) until a queue starts to form on route $R(t') + 1$ or the queue vanishes at $R(t')$,*
- Step 3.** *else if $\Delta(t') = \Delta_{R(t)-1}^*$ then set $R(t') = R(t) - 1$,*
- Step 4.** *if the queue vanishes in the freeway stop, else update $t = t'$ and go to Step 1.*

The simplified graphical solution method of §3.1 can be streamlined using the following algorithm:

Algorithm 2 *Set $t = t_0 = 0, R(t_0) = 0, M = \mu_0$ and $D(s) = sM, s \geq 0$. Then perform the following steps:*

- Step 1.** *identify k and the earliest time $t' \geq t$ such that: $\Delta(t') = \Delta_{R(t)+k}^*$; $k = -1, 0, 1$.*
- Step 2.** *set $R(t') = R(t) + k, M = M + \mu_{R(t')}$,*
- Step 3.** *if $R(t') = 0$ then stop, else set $D(s) = D(t' + \Delta(t')) + (s - t')M, s \geq t'$, update $t = t'$ and go to Step 1.*

B Logistic smoothing

This appendix shows how to approximate a bilinear arrival curve $A(t)$ that violates the stability condition (20) by a smooth arrival curve $\tilde{A}(t)$ using logistic smoothing. The idea is to smooth out the transition between normalized flow levels c_1 and c_2 at T_1 of $A(t)$ by using:

$$\dot{\tilde{A}}(t)/\mu_0 = \frac{c_1 - c_2}{1 + e^{-b(t-T_1)}} + c_2, \quad (\text{B.1})$$

where $b > 0$ controls how fast the transition takes place ($b = 20$ was used in Fig. 6). The corresponding equilibrium travel time is given by

$$\tilde{\Delta}(t) = \frac{c_1 - 1}{a}(1 - e^{-at}) + \frac{c_1 - c_2}{a + b}F(t), \quad (\text{B.2})$$

where $F(t) = e^{-at-bT_1} \mathfrak{N}(-e^{-bT_1}) - e^{b(t-T_1)} \mathfrak{N}(-e^{b(t-T_1)})$ and $\mathfrak{N}(\cdot)$ gives Gauss Hypergeometric function with parameters $(1, (a+b)/b, a/b+2)$. This function has the series expansion $\mathfrak{N}(z; \alpha, \beta, \gamma) = \sum_{k=0}^{\infty} \frac{z^k (\alpha)_k (\beta)_k}{k! (\gamma)_k}$ where $(\alpha)_k = \prod_{i=0}^{k-1} (i + \alpha)$. It is worth noting that although $\mathfrak{N}(\cdot)$ has to be evaluated numerically in general, in cases where $a = b$ or $2a = b$ this function is algebraic, which makes $F(t)$ and thus $\tilde{\Delta}(t)$ algebraic, i.e.:

$$F(t) = \begin{cases} 2(1 - e^{-at} + e^{-a(t-T_1)} \log \frac{1+e^{a(t-T_1)}}{1+e^{-aT_1}}) & \text{if } a = b, \\ 3(1 - e^{-at} + e^{-a(t-T_1)} (\tan^{-1}(e^{a(t-T_1)}) - \cot^{-1}(e^{aT_1}))) & \text{if } 2a = b. \end{cases} \quad (\text{B.3})$$