Mathematical Programming
https://doi.org/10.1007/s10107-021-01725-z

FULL LENGTH PAPER

Series B

Algorithms for flows over time with scheduling costs

Dario Frascaria[1] · **Neil Olver**[2,3]

Received: 30 June 2020 / Accepted: 17 October 2021

Abstract
Flows over time have received substantial attention from both an optimization and (more recently) a game-theoretic perspective. In this model, each arc has an associated delay for traversing the arc, and a bound on the rate of flow entering the arc; flows are time-varying. We consider a setting which is very standard within the transportation economic literature, but has received little attention from an algorithmic perspective. The flow consists of users who are able to choose their route but also their departure time, and who desire to arrive at their destination at a particular time, incurring a *scheduling cost* if they arrive earlier or later. The total cost of a user is then a combination of the time they spend commuting, and the scheduling cost they incur. We present a combinatorial algorithm for the natural optimization problem, that of minimizing the average total cost of all users (i.e., maximizing the social welfare). Based on this, we also show how to set tolls so that this optimal flow is induced as an equilibrium of the underlying game.

Keywords Flows over time · Tolls · Traffic

Mathematics Subject Classification Primary: 90C27 · Secondary: 05C21 · 90B20

Partially supported by NWO TOP Grant 614.001.510 and NWO Vidi Grant 016.Vidi.189.087.
A preliminary version of this paper has appeared in proceedings of the the 21st Conference on Integer Programming and Combinatorial Optimization (IPCO), 2020.

✉ Dario Frascaria
d.frascaria@gmail.com

Neil Olver
N.Olver@lse.ac.uk

1 Department of Econometrics and Operations Research, Vrije Universiteit Amsterdam, Amsterdam, The Netherlands

2 Department of Mathematics, London School of Economics and Political Science, London, UK

3 CWI, Amsterdam, The Netherlands

Published online: 01 November 2021

1 Introduction

The study of *flows over time* is a classical one in combinatorial optimization; it began already with the work of Ford and Fulkerson [10] in the 50s. It is a natural extension of static flows, which associates a single numerical value, representing a total quantity or rate of flow on the arc. In a flow over time, a second value associated with each arc represents the time it takes for flow to traverse it; the flow is then described by a function on each arc, representing the rate of flow entering the arc as a function of time.

Classical optimization problems involving static flows have natural analogs in the flow over time setting (see the surveys [17,26]). For example (restricting the discussion to single commodity flows), the *maximum flow over time* problem asks to send as much flow as possible, departing from the source starting from time 0 and arriving to the sink by a given time horizon T; this can be solved in polynomial time [9–11]. A *quickest flow* asks, conversely, for the shortest time horizon necessary to send a given amount of flow. Of particular importance for us is the notion of an *earliest arrival flow*: this has the very strong property that simultaneously for all $T' \leq T$, the amount of flow arriving by time T' is as large as possible [13]. Such a flow can also be characterized as minimizing the average arrival time [15]. Earliest arrival flows can be "complicated", in that they can require exponential space (in the input size) to describe [31], and determining the average arrival time of an earliest arrival flow is NP-hard [8]. But they can be constructed in time strongly polynomial in the sum of the input and output size [3].

Another important aspect of many settings where flow-over-time models are applicable—such as traffic—involves game theoretic considerations. In traffic settings, the flow is made up of a large number of individuals making their own routing choices, and aiming to maximize their own utility rather than the overall social welfare (e.g., average journey time). *Dynamic equilibria*, which is the flow over time equivalent of Wardrop equilibria for static flows, are key objects of study. Existence, uniqueness, structural and algorithmic issues, and much more have been receiving increasing recent interest from the optimization community [4–7,16,23,24].

Traffic, being such a relevant and important topic, has received attention from many different communities, each with their own perspective. Within the transportation economic literature, modelling other aspects of user choice besides route choice has been considered particularly important. A very standard setting, motivated by morning rush-hour traffic, is the following [2,28]. Users are able to choose not only their route, but also their *departure time*. They are then concerned not only with their journey time, but also their *arrival time* at the destination. This is captured in a *scheduling cost function* which we will denote by ρ: a user arriving at time θ will experience a scheduling cost of $\rho(\theta)$. The total disutility of a user is then the sum of their scheduling cost and their journey time (scaled by some factor $\alpha > 0$ representing their value for time spent commuting). A very standard choice of ρ is

$$\rho(\theta) = \begin{cases} -\beta\theta & \text{if } \theta \leq 0 \\ \gamma\theta & \text{if } \theta > 0 \end{cases}, \tag{1}$$

where 0 is the desired arrival time and $0 < \beta < \alpha < \gamma$ (it is very bad to be late, but time spent in the office early is better than time spent in traffic). We will allow general scheduling cost functions, though for most of the paper we will focus on strongly unimodal cost functions; these are the most relevant, and this avoids some distracting technical details.

Two very natural questions can be posed at this point. The first is a purely optimization question, with no attention paid to the decentralized nature of traffic.

Problem 1 How can one compute a flow over time minimizing the average total cost paid by users, i.e., maximizing the social welfare?

From now on, we will call a solution to this problem simply an *optimal flow*.

It is well understood that users will typically not coordinate their actions to induce a flow that minimizes total disutility. There is a huge body of literature (particularly in the setting of static flows [19]) investigating this phenomenon. In the traffic setting, the relevance of an optimal flow represented by an answer to this question comes primarily via the possibility of *pricing*. By putting appropriate tolls on roads, we can influence the behaviour of users and the resulting dynamic equilibrium. Thus:

Problem 2 How can one set tolls (possibly time-varying) on the arcs of a given instance so that an optimal flow is obtained in dynamic equilibrium?

One subtlety is that since dynamic equilibria need not be precisely unique, there is a distinction between tolls that induce an optimal flow as *an* equilibrium, compared to tolls for which *all* dynamic equilibria are optimal. We will call this *weak* and *strong* enforcement of optimality, respectively, and will return to this subtlety shortly. (See Harks [14] for some related notions of enforcement in a general pricing setting.)

It is also natural to ask about equilibria *without* tolls. This is closely related to work on dynamic equilibria under *exogenous* demand—meaning that users do not have a choice of departure time, but enter the network according to some given rate function. Issues of existence and uniqueness, algorithmic concerns, and other properties of dynamic equilibria in this setting, have received a lot of attention (see above for a list of relevant references). However, the modelling issues are rather orthogonal to the thrust of this paper, so we will not discuss this in any more detail. We refer the interested reader to Frascaria, Olver and Verhoef [12], who define dynamic equilibria in the endogenous departure model and investigate some properties of it.

Questions like these are of great interest to transportation economists. However, most work in that community has focused on obtaining a fine-grained understanding of very restricted topologies (such as a single link, or multiple parallel links); see [27] for a survey.

Both of these questions (for general network topologies) were considered by Yang and Meng [30] in a discrete time setting, by exploiting the notion of *time-expanded graphs*. This is a standard tool in the area of flows over time; discrete versions all of the optimization questions concerning flows over time mentioned earlier can (in a sense) be dealt with in this way. A node v in the graph is expanded to a collection (v, i) of nodes, for $i \in \mathbb{Z}$ in a suitable interval, and an arc vw of delay τ_{vw} becomes a collection of arcs $((v, i), (w, i+\tau_{vw}))$ (this assumes a scaling so that τ_{vw} is a length in multiples of

the chosen discrete timesteps). Scheduling costs are encoded by appropriately setting arc costs from (t, i) to a supersink t' for each i, and the problem can be solved by a minimum cost static flow computation. A primary disadvantage of this approach (and in the use of time-expanded graphs more generally) is that the running time of the algorithm depends polynomially on the number of time steps, which can be very large. Further, it cannot be used to exactly solve the continuous time version (our interest in this paper); by discretizing time, it can be used to approximate it, but the size of the time-expanded graph is inversely proportional to the step size of the discretization. In the same work [30], the authors also observe that in the discrete setting, an answer to the second question can be obtained from the time-expanded graph as well. Taking the LP describing the minimum cost flow problem on the time-expanded graph, the optimal dual solution to this LP provides the necessary tolls to enforce (weakly) an optimal flow. (This is no big surprise—dual variables can frequently be interpreted as prices.)

An assumption on ρ Suppose we consider ρ in the standard form given in (1), but with $\beta > \alpha$. This means that commuting is considered to be less unpleasant than arriving early. A user arriving earlier than time 0 at the sink would be better off "waiting" at the sink before leaving, in order to pay a scheduling cost of 0. Whether waiting in this way is allowed or not depends on the precise way one specifies the model, but it is most natural (and convenient) to allow this. If we do so, then it is clear that a scheduling cost function ρ can be replaced by

$$\hat{\rho}(\theta) := \min_{\xi \geq \theta} \rho(\xi) + \alpha(\xi - \theta)$$

without changing the optimal flow (except there is no longer any incentive to wait at the sink, and we need not even allow it). Then $\theta \to \hat{\rho}(\theta) + \alpha\theta$ is nondecreasing. From now on, we always assume that ρ satisfies this; we will call it the *growth bound* on ρ.

Our results We give a combinatorial algorithm to compute an optimal flow. Similarly to the case of earliest arrival flows, this flow can be necessarily complicated, and involves a description length that is exponential in the input size.

The algorithm is also similar to that for computing an earliest arrival flow. It is based on the (possibly exponentially sized) path decomposition of a minimum cost flow into *successive shortest paths*. In particular, suppose we choose the scheduling cost function to be

$$\rho(\theta) = \begin{cases} -\alpha\theta & \text{if } \theta \leq 0 \\ \infty & \text{if } \theta > 0 \end{cases}. \tag{2}$$

Then the disutility a user experiences is precisely described by how much before time 0 they depart; all users must arrive by time 0 to ensure finite cost. This is precisely the reversal (both in time and direction of all arcs) of an earliest arrival flow, from the sink to the source. (By writing the average arrival time objective as the integral over time of the total flow not yet arrived by this time, this exact correspondence is easy to see.) Our algorithm, in this case, is the same (up to the time reversal) as the usual algorithm for earliest arrival flow [9].

This also shows that there are instances where all optimal solutions to Problem 1 require exponential size (as a function of the input encoding length), since this is the case for earliest arrival flows [31].

Despite the close relation to earliest arrival flows, the proof of optimality of our algorithm is rather different. A key reason for this is the following. As mentioned, earliest arrival flows have the strong property that the amount of flow arriving before a given deadline T' is the maximum possible, *simultaneously for all choices of* T' (up to some maximum depending on the total amount of flow being sent). This implies that an earliest arrival flow certainly minimizes the average arrival time amongst all possible flows [15], but is a substantially stronger property. A natural analog of this stronger property in our setting would be to ask for a flow for which, simultaneously for any given cost horizon $C' \leq C$, the amount of flow consisting of agents experiencing disutility at most C' is as large as possible. Unfortunately, in general no such flow exists. The example is too involved to discuss here, but it relates to some questions on the behaviour of dynamic equilibria in this model that are investigated in a parallel manuscript [12].

Since the proofs for earliest arrival flows [3,13,18,29] show this stronger property which does not generalize, we take a different approach. Our proof is based on duality (of an infinite dimensional LP, though we do not require any technical results on such LPs). The main technical challenge in our work comes from determining the correct ansatz for the dual solution, as well as exploiting properties of the residual networks obtained from the successive shortest paths algorithm in precisely the right way to demonstrate certain complementary slackness conditions.

We remark that some of the work on maximum flow over time does make the connection to infinite dimensional LPs; see Sharkey [25] for a survey and some further references. In particular, we point out the flow-over-time version of max-flow min-cut by Philpott [20], which can be viewed as a derivation of strong duality for the corresponding infinite linear program.

As was the case with the time-expanded graph approach, the optimal dual solution immediately provides us with corresponding tolls for which the optimal flow is an equilibrium. However, we obtain an explicit formula for the optimal tolls, in terms of the successive shortest paths of the graph (see Sect. 3). This may be useful in obtaining a better structural understanding of optimal tolls, beyond just their computation. We also remark that a corollary of our result is that there is always an optimal solution without waiting (except at the source).

Consider for a moment the model where users cannot choose their departure time, but instead are released from the source at a fixed rate u_0, and simply wish to reach the destination as early as possible. This is the game-theoretic model that has received the most attention from the flow-over-time perspective [4,6,7,16,24]. Our construction of optimal tolls is applicable to this model as well, as discussed in Sect. 5. As far as we are aware, no explicit description of optimal tolls was previously known even in this setting.

We now return to the subtlety alluded to earlier: the distinction between strongly enforcing an optimal flow, and only weakly enforcing it.

Consider the simple instance in Fig. 1. Suppose that the outflow of arc a is larger than 1 for some period in the optimum flow, due to the choice of scheduling cost

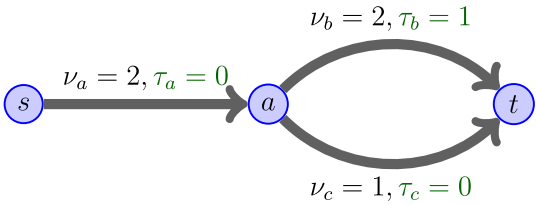


Fig. 1 An instance where time-varying arc tolls cannot enforce that *all* equilibria are optimal flows

function. In this period, one unit of flow would take the bottom arc c, and the rest will be routed on b. Since the total cost (including tolls) of all users is the same in a tolled dynamic equilibrium, a toll of cost equivalent to a unit delay on arc c is needed in this period to induce the optimal flow. But then it will also be an equilibrium to send *all* flow in this period along b.

To strongly enforce an optimal flow, we need more flexible tolls. One way that we can do it is by "tolling lanes". If we are allowed to dynamically divide up the capacity of an arc into "lanes" (say a "fast lane" and a "slow lane"), and then separately set time-varying tolls on each lane, then we *can* strongly enforce any optimal flow. We discuss this further in Sect. 5. We are not aware of settings where this phenomenon has been previously observed, and it would be interesting to explore this further in a more applied context.

Outline of the paper We introduce some basic notation and notions, as well as formally define our model, in Sect. 2. In Sect. 3, we describe our algorithm, and show that it returns a feasible flow over time; we restrict ourselves to the most relevant case of a strictly unimodal scheduling cost function. In Sect. 4 we show optimality of this algorithm, and in Sect. 5 we derive optimal tolls from this analysis. Finally, in Sect. 6 we discuss general scheduling cost functions.

2 Model and preliminaries

We use $[n]$ to denote the set $\{1, 2, \ldots, n\}$, for any positive integer n, and $\mathbb{R}_+$ to denote the nonnegative reals. The notation $(z)^+$ is used to denote the nonnegative part of z, i.e., $(z)^+ := \max\{z, 0\}$. Given $v \in \mathbb{R}^X$ and $A \subseteq X$, we will use the shorthand notation $v(A) := \sum_{a \in A} v_a$. All graphs considered will be directed. We assume all graphs to be simple, and that there are no digons (i.e., there are no pairs $v, w \in V$ so that vw and wv are both arcs). This is for notational convenience only—this restriction can easily be lifted.

We begin with some basic notions and results about static flows and flows over time. For further details regarding static flows, we refer the reader to the book by Ahuja, Magnanti and Orlin [1]. For more about flows over time, we suggest the surveys by Skutella [26] and Köhler et al. [17].

Static flows Let $G = (V, E)$ be a directed graph, with source node $s \in V$ and sink node $t \in V$. Each arc $e \in E$ has a *capacity* ν_e and a *delay* τ_e (both nonnegative). We use $\delta^+(v)$ to denote the set of arcs in E with tail v, and $\delta^-(v)$ the set of arcs with head v.

Consider some $f \in \mathbb{R}^E$. For $v \in V$, we define the *net flow* at v (denoted ∇f_v) to be the quantity

$$\nabla f_v := f(\delta^-(v)) - f(\delta^+(v)) = \sum_{e \in \delta^-(v)} f_e - \sum_{e \in \delta^+(v)} f_e.$$

We say that f is a *(static) s-t-flow of value Q* if

(i) $\nabla f_v = 0$ for all $v \in V \setminus \{s, t\}$, with $\nabla f_t = -\nabla f_s = Q$; and
(ii) $0 \leq f_e \leq \nu_e$ for all $e \in E$.

Given an s-t-flow f, its *residual network* $G^f = (V, E^f)$ is defined by

$$E^f = \{vw : vw \in E \text{ and } f_{vw} < \nu_{vw}\} \cup \{vw : wv \in E \text{ and } f_{wv} > 0\}.$$

Call arcs in $E^f \cap E$ *forward arcs* and arcs in $E^f \setminus E$ *backwards arcs*. The *residual capacity* ν_e^f of an arc $e \in E^f$ is then $\nu_{vw}^f := \nu_{vw} - f_{vw}$ for vw a forward arc, and $\nu_{vw}^f := f_{wv}$ for vw a backwards arc. We also define $\tau_{vw} := -\tau_{wv}$ for all backwards arcs vw.

Given a subset F of arcs, we use $\chi(F)$ to denote the characteristic vector of F. In particular, if P is a path from v to w, then $\chi(P)$ is a unit flow from v to w.

We make the definitions $\overleftarrow{E} := \{wv : vw \in E\}$ and $\overrightarrow{E} := E \cup \overleftarrow{E}$. We will regard a vector $g \in \mathbb{R}_+^{\overrightarrow{E}}$ as a flow in $(V, \overrightarrow{E})$ if for every $vw \in E$, either $g_{vw} = 0$ or $g_{wv} = 0$. Given two such flows f and g, we define their sum $f + g$ by taking the sum as vectors, and then cancelling flows on oppositely directed arcs if necessary (so $(f + g)_{vw}$ and $(f + g)_{wv}$ are never both nonzero). Define $f - g$ similarly.

Given a choice of value Q, a *minimum cost flow* is an s-t-flow f^* minimizing $\sum_{e \in E} f_e \tau_e$ (amongst all s-t-flows f of value Q). An s-t-flow f (of the correct value) is a minimum cost flow if and only if E^f contains no negative cost cycles, i.e., cycles $C \subseteq E^f$ with $\tau(C) < 0$.

Flows over time Let $\mathcal{L}$ denote the space of measurable functions on $\mathbb{R}$ with compact support. Consider some vector $f \in \mathcal{L}^E$. Define the *net flow into v at time θ* by

$$\nabla f_v(\theta) := \sum_{e \in \delta^-(v)} f_e(\theta - \tau_e) - \sum_{e \in \delta^+(v)} f_e(\theta).$$

Note that $f_e(\theta)$ represents the flow *entering* arc e at time θ; this flow will exit the arc at time $\theta + \tau_e$ (explaining the asymmetry between the terms for flow entering and flow leaving in the above).

We say that $f \in \mathcal{L}^E$ is a *flow over time of value Q* if the following hold.

(i) $\int_{-\infty}^{\infty} \nabla f_v(\theta) d\theta = Q(\mathbf{1}_{v=t} - \mathbf{1}_{v=s})$ for all $v \in V$.
(ii) $\int_{-\infty}^{\xi} \nabla f_v(\theta) d\theta \geq 0$ for all $v \in V \setminus \{s\}$ and $\xi \in \mathbb{R}$.
(iii) $0 \leq f_e(\theta) \leq \nu_e$ for all $e \in E$ and $\theta \in \mathbb{R}$.

Note that this definition allows for flow to wait at a node; to disallow this and consider only *flows over time without waiting*, we would additionally require with the condition

(iv) $\nabla f_v(\theta) = 0$ for all $v \in V \setminus \{s, t\}$ and $\theta \in \mathbb{R}$.

We also have a natural notion of a residual network in the flow over time setting. Define, for any flow over time f and $\theta \in \mathbb{R}$,

$$E^f(\theta) = \{vw : vw \in E \text{ and } f_{vw}(\theta) < \nu_{vw}\} \cup \{vw : wv \in E \text{ and } f_{wv}(\theta - \tau_{wv}) > 0\}.$$

Minimizing scheduling cost We are concerned with the following optimization problem. Given a *scheduling cost function* $\rho : \mathbb{R} \to \mathbb{R}_+$, as well as a value $\alpha > 0$, determine a flow over time f of value Q that minimizes the sum of the *commute cost* $\alpha \sum_{e \in E} \tau_e \cdot \int_{-\infty}^{\infty} f_e(\theta) d\theta$ and the *scheduling cost* $\int_{-\infty}^{\infty} \nabla f_t(\theta) \cdot \rho(\theta) d\theta$. As already discussed, we assume that ρ satisfies the growth bound, i.e., that $\theta \to \rho(\theta) + \alpha\theta$ is nondecreasing. This ensures that waiting at t is not needed, which is in fact disallowed by our definition,[1] and makes various arguments cleaner. We will also make the assumption that ρ is strongly unimodal.[2] We then assume w.l.o.g. that the minimizer of ρ is at 0, and that $\rho(0) = 0$. For further technical convenience, by adjusting ρ on a set of measure zero we take ρ to be lower semi-continuous.[3]

The above conditions will suffice for our structural characterization of an optimum flow and its analysis, but more is needed in order to be able to implement the algorithm. The algorithm will require not just oracle access to ρ, but also to ρ^{-1}. That is, given $y > 0$, we are able to query the pair of solutions (one positive, one negative) that map to y under ρ. In order to ensure that the optimal solution has a rational description, we should require not only that ρ maps rationals to rationals, but that ρ^{-1} does too; a simple function like $\rho(\theta) = \theta^2$ that violates this can lead to irrational optimum solutions, as we will remark on later. For algorithmic purposes, it is sensible to restrict attention to scheduling costs that are explicitly given as piecewise linear functions; we will focus primarily on this case.

The assumption of strong unimodality is not necessary; the algorithm and analysis can be extended (with some additional effort). We postpone this discussion to the end of the paper.

3 A combinatorial algorithm

In this section we present an algorithm that computes an optimal flow over time, assuming that ρ is strongly unimodal. The proof of optimality is discussed in Sect. 4.

3.1 Successive shortest paths

We begin by recalling the *successive shortest paths (SSP)* algorithm for computing a minimum cost static flow. It is not a polynomial time algorithm, so it is inefficient as an algorithm for static flows, but it provides a structure that is relevant for flows over

[1] Were this really needed, one could simply add a dummy arc tt' to a new sink t'.

[2] I.e., strictly decreasing until some moment, and then strictly increasing.

[3] Since an increasing function is continuous almost everywhere, we can replace $\rho(\theta)$ by $\lim_{\epsilon \downarrow 0} \rho(\theta + \epsilon)$ for all $\theta \geq 0$; and similarly with $\lim_{\epsilon \uparrow 0} \rho(\theta + \epsilon)$ for $\theta < 0$.

time. This is of course well known from its role in constructing earliest arrival flows, which we will briefly detail.

The SSP algorithm construct a sequence of paths $(P_1, P_2, \ldots)$ and associated amounts $(x_1, x_2, \ldots)$ inductively as follows. Suppose $P_1, \ldots, P_j$ and $x_1, \ldots, x_j$ have been defined. Let

$$f^{(j)} = \sum_{i=1}^{j} x_i \chi(P_i),$$

and let G_j denote the residual graph of $f^{(j)}$ (G_0 being the original network). Also let $d_j(v, w)$ denote the length (w.r.t. arc delays τ in G_j) of a shortest path from v to w in G_j (this may be infinite). By construction, G_j will contain no negative cost cycles, so that d_j is computable. If $d_j(s, t) = \infty$, we are done; set $m := j$. Otherwise, define P_{j+1} to be any shortest s-t-path in G_j, and x_{j+1} the minimum capacity in G_j of an arc in P_{j+1}. It can be shown that $\sum_{j=1}^{r} \tilde{x}_j \chi(P_j)$, with r and $\tilde{x}$ defined such that $\tilde{x}_j = x_j$ for $j < r$, $0 \leq \tilde{x}_r \leq x_r$ and $\sum_{j=1}^{r} \tilde{x}_j = M$, is a minimum cost flow of value M, as long as M is not larger than the value of a maximum flow.

To construct an earliest arrival flow with time horizon T, we (informally) send flow at rate x_j along path P_j for the time interval $[0, T - \tau(P_j)]$, for each $j \in [m]$ (if $\tau(P_j) > T$, we send no flow along the path). By this, we mean that for each $e = vw \in P_j$, we increase by x_j the value of $f_e(\theta)$ for $\theta \in [d_{j-1}(s, v), T - d_{j-1}(v, t)]$ (or if e is a backwards arc, we instead decrease $f_{wv}(\theta - \tau_{wv})$). An argument is needed to show that this defines a valid flow, since we must not violate the capacity constraints, and moreover, P_j may contain reverse arcs not present in G (see, e.g., [26]).

3.2 The algorithm

We are now ready to describe our algorithm for minimizing the disutility, which is a natural variation on the earliest arrival flow algorithm. It is also constructed from the successive shortest paths, but using a *cost horizon* rather than a *time horizon*. For now, consider C to be a given value (it will be the "cost horizon").

- For each $j \in [m]$, let $[a_j, b_j]$ be the maximal interval so that

$$[\rho(\xi + d_{j-1}(s, t)) \leq C - \alpha d_{j-1}(s, t) \quad \text{for all} \quad \xi \in [a_j, b_j].$$

  (If ρ is continuous, then of course $\rho(a_j + d_{j-1}(s, t)) = \rho(b_j + d_{j-1}(s, t)) = C - \alpha d_{j-1}(s, t)$). Note that a user leaving at time a_j or b_j and using path P_j, without waiting at any moment, incurs disutility C; whereas a user leaving at some time $\theta \in (a_j, b_j)$ and using path P_j will incur a strictly smaller total cost.
- Construct the flow over time which sends flow at rate x_j along path P_j for the time interval $[a_j, b_j]$, for each $j \in [m]$. More formally: for any $j \in [m]$ and node $v \in P$, let $\tau_j(v)$ denote the sum of the delays on the arcs of P_j between v and t.

Then the algorithm returns the vector $f \in \mathcal{L}^E$ given by

$$f_{vw}(\theta) = \sum_{\substack{j\in[m]:vw\in P_j \\ \theta+\tau_j(v)\in[a_j,b_j]}} x_j \quad - \quad \sum_{\substack{j\in[m]:wv\in P_j \\ \theta+\tau_j(v)\in[a_j,b_j]}} x_j \quad \text{for all} \quad vw \in E, \theta \in \mathbb{R}.$$

We remark that this structure is closely related to that of "generalized temporally repeated flows" [26].

As we will shortly argue, the vector f returned by this algorithm is a feasible flow over time (the main issue is checking that f is nonnegative and satisfies the capacity constraints). Given this, its value will be $\sum_{j=1}^{m} x_j(b_j - a_j)$. Since ρ is strongly unimodal, this value changes continuously and monotonically with C. Thus a bisection search can be used to determine the correct choice of C for a given value Q, at least to within some predetermined error ϵ. Determining the *precisely* correct value of C may not be possible without some additional information about ρ.

If ρ is piecewise linear (as is the case, in particular, for the "standard" β/γ choice generally used in the transportation economics literature), bisection search can be avoided. Let K_l and K_r denote the number of linear segments of ρ to the left and right of 0, respectively, and let $K = K_l + K_r$. Write $a_j(C)$ and $b_j(C)$ to explicitly indicate the dependence of the interval in which flow is sent along P_j as a function of C. Then let $Q_j(C) := x_j(b_j(C) - a_j(C))$; this is the total mass sent along path P_j in the solution obtained with time horizon C. Now notice that $a_j(C)$ is a piecewise linear function with at most $K_l + 1$ linear segments; it is defined by $\rho(a_j(C) + d_{j-1}(s,t)) = C - \alpha d_{j-1}(s,t)$ for $C \geq \alpha d_{j-1}(s,t)$, and $a_j(C) = -d_{j-1}(s,t)$ otherwise. Similarly, $b_j(C)$ is a piecewise linear function with at most $K_r + 1$ linear segments. The total value $\sum_{j=1}^{m} Q_j(C)$ is thus piecewise linear with at most $m(K+2)$ linear segments. Thus, even the entire parametric curve of cost horizon C against flow value Q can be computed in time $O(mK)$, once the successive shortest paths have been computed.

Before proving the correctness of our algorithm, we show an example of a flow over time minimizing a given scheduling cost, as would be constructed by our algorithm.

Example 1 Consider the graph $G = (V, E)$ illustrated in Fig. 2a, with $V = \{s, a, b, t\}$, $E = \{sa, sb, ab, at, bt\}$ and capacities ν_e and delays τ_e as indicated in the figure, in the order (ν_e, τ_e).

The successive shortest paths are $P_1 = \{s, a, b, t\}$, with length 3, $P_2 = \{s, a, t\}$ and $P_3 = \{s, b, t\}$, both with length 4, and $P_4 = \{s, b, a, t\}$ with length 5. All the associated amounts are equal to 1 (see Fig. 2b–d).

Consider now a cost horizon C equal to 6 and the standard scheduling cost function given in Equation 1 with $\alpha = 1$, $\beta = 0.5$, $\gamma = 2$. Our algorithm then sends 1 unit of flow along P_1 for the time interval $[-9, -1.5]$; 1 unit of flow along P_2 and 1 along P_3 for the time interval $[-8, -3]$; and 1 unit of flow along P_4 for the time interval $[-7, -4.5]$. The resulting flow is described in Fig. 3 with a sequence of snapshots of the network.

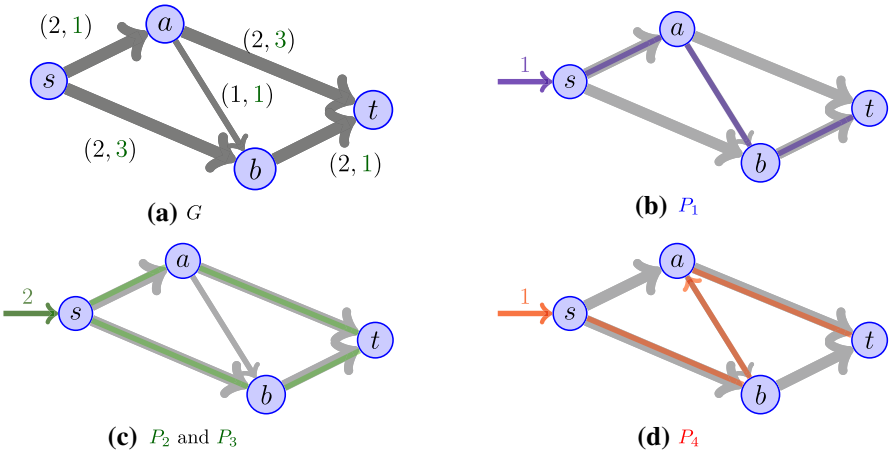


Fig. 2 The network and the successive shortest paths of Example 1

3.3 Feasibility

In the following we show that the resulting flow f is a feasible flow over time. Given a vertex $v \in V$, a time $\theta \in \mathbb{R}$ and $j \in [m]$, let

$$c_j(v, \theta) = \alpha d_{j-1}(s, t) + \rho(\theta + d_{j-1}(v, t)).$$

If $v \in P_j$ then $c_j(v, \theta)$ is the total cost of a user that utilizes path P_j and passes through node v at time θ; there does not seem to be a simple interpretation if $v \notin P_j$ however. Now define

$$J(v, \theta) = \max\{j \in [m] : c_j(v, \theta) \leq C\}, \tag{3}$$

with the convention that the maximum over the empty set is 0. We remark for future reference that since $d_m(s, t) = \infty$, we do have that

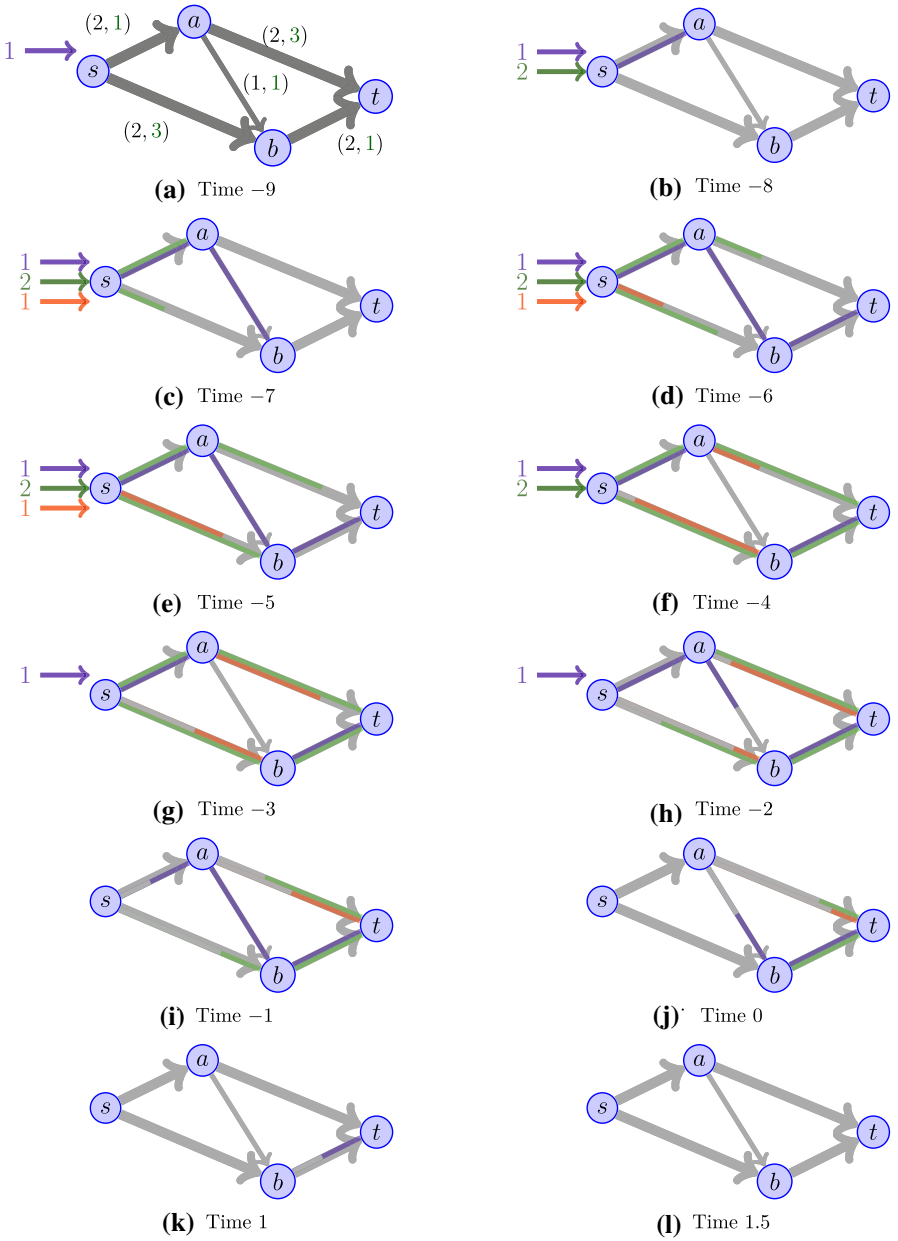
$$\alpha d_{J(v,\theta)}(s, t) + \rho(\theta + d_{J(v,\theta)}(v, t)) > C. \tag{4}$$

The motivation for this definition comes from the following theorem, which completely characterizes f in terms of the static flows arising from successive shortest paths. (If preferred, one could even think of this theorem as providing the definition of f.)

Theorem 1 *$f_{vw}(\theta) = f_{vw}^{(J(v,\theta))}$ for any $vw \in E$ and $\theta \in \mathbb{R}$.*

Before proving Theorem 1, we need the following lemma.

Lemma 2 *$c_j(v, \theta)$ is nondecreasing in j for any $\theta \in \mathbb{R}$ and $v \in V$.*



(a) Time −9 **(b)** Time −8

(c) Time −7 **(d)** Time −6

(e) Time −5 **(f)** Time −4

(g) Time −3 **(h)** Time −2

(i) Time −1 **(j)** Time 0

(k) Time 1 **(l)** Time 1.5

Fig. 3 Chronological sequence of snapshots of the optimal flow for the instance given in Example 1

Proof Consider any $j \in [m-1]$; we show that $c_{j+1}(v,\theta) \geq c_j(v,\theta)$. Suppose R is a shortest v-t-path in G_{j-1}, so $\tau(R) = d_{j-1}(v,t)$. Consider the unit v-t flow $g = \chi(P_{j+1}) - \chi(P_j) + \chi(R)$ in $\overleftrightarrow{E}$. Now observe that the support of g is contained in G_j: P_{j+1} and $\overleftarrow{P_j}$ are certainly contained in G_j; and if $e \in R \cap (E_{j-1} \setminus E_j)$, then $e \in P_j$, which means $g_e = 0$. Since G_j contains no negative cost cycles, the cost of g is at least that of a shortest v-t-path in G_j, and so

$$d_j(v,t) \leq \tau(P_{j+1}) - \tau(P_j) + \tau(R) = d_j(s,t) - d_{j-1}(s,t) + d_{j-1}(v,t)\,.$$

Finally, we can conclude

$$\begin{aligned}\alpha d_j(s,t) + \rho(\theta + d_j(v,t)) &= \alpha d_j(s,t) + \rho(\theta + d_{j-1}(v,t)) - \rho(\theta + d_{j-1}(v,t)) + \rho(\theta + d_j(v,t)) \\ &\geq \alpha d_j(s,t) + \rho(\theta + d_{j-1}(v,t)) - \alpha(d_j(v,t) - d_{j-1}(v,t)) \\ &\geq \alpha d_{j-1}(s,t) + \rho(\theta + d_{j-1}(v,t)),\end{aligned}$$

where the first inequality follows from the growth bound, using $d_j(v,t) \geq d_{j-1}(v,t)$. □

Proof of Theorem 1 Fix some $vw \in E$ and $\theta \in \mathbb{R}$. Consider now any $j \in [m]$ for which $\alpha\tau(P_j) \leq C$ (so that P_j is used for a nonempty interval) and $vw \in P_j$. Since P_j is a shortest path in G_{j-1}, if we send flow along this path starting from some time ξ, it will arrive at v at time $\xi + d_{j-1}(s,v)$. Considering the definition of the interval $[a_j, b_j]$, we see that P_j contributes flow to vw at time θ if $c_j(v,\theta) \leq C$. By Lemma 2, this occurs precisely if $j \leq J(v,\theta)$.

Considering in similar fashion paths P_j with $wv \in P_j$ (and noting that $J(w, \theta + \tau_{vw}) = J(v,\theta)$), we determine that

$$f_{vw}(\theta) = \sum_{\substack{j: vw \in P_j \\ j \leq J(v,\theta)}} x_j - \sum_{\substack{j: wv \in P_j \\ j \leq J(v,\theta)}} x_j = f_{vw}^{(J(v,\theta))}.$$

□

Feasibility of f is now immediate.

Corollary 3 *f is a feasible flow over time without waiting.*

Proof By the way that we constructed f, it has value Q, satisfies flow conservation, and has no waiting. Only nonnegativity and the capacity constraint remain, which follows from Theorem 1. □

4 Optimality

In this section, we show that our proposed algorithm does return an optimal flow.

4.1 Duality-based certificates of optimality

We can write the problem we are interested in as an infinite continuous linear program as follows:

$$
\begin{aligned}
\min \quad & \int_{-\infty}^{\infty} \rho(\theta) \nabla f_t(\theta) d\theta + \alpha \sum_{e \in E} \tau_e \int_{-\infty}^{\infty} f_e(\theta) d\theta + \alpha \sum_{v \in V \setminus \{s,t\}} \int_{-\infty}^{\infty} z_v(\theta) d\theta \\
\text{s.t.} \quad & \int_{-\infty}^{\infty} \nabla f_s(\theta) d\theta = -Q \\
& \int_{-\infty}^{\infty} \nabla f_t(\theta) d\theta = Q \\
& \int_{-\infty}^{\theta} \nabla f_v(\xi) d\xi = z_v(\theta) \qquad \forall v \in V \setminus \{s,t\}, \theta \in \mathbb{R} \\
& f_e(\theta) \leq \nu_e \qquad \forall e \in E, \theta \in \mathbb{R} \\
& z, f \geq 0
\end{aligned}
\tag{5}
$$

Here, $z_v(\theta)$ represents the amount of flow waiting at node v at time θ (which must always be nonnegative). Both f_e for any $e \in E$ and z_v for any $v \in V$ should be bounded and measurable functions with compact support. This implies that in fact z_v is absolutely continuous for each $v \in V$. Note that the objective function captures separately the contribution to the journey time coming from actually travelling across arcs, and the contribution from waiting at nodes. As a further remark, this linear program does not explicitly prevent flow from departing and then returning to t; only the aggregate constraint $\int_{-\infty}^{\infty} \nabla f_t(\theta) = Q$ is imposed. The growth condition on ρ, however, ensures that it is never profitable to do this; the reduction of scheduling cost is never more than the cost incurred in travelling (including waiting at nodes).

Given that this is an infinite-dimensional linear program, one may reasonably expect to be able to write down a dual, and make use of weak and strong duality, as well as complementary slackness conditions. However, care is needed: the situation for infinite (even countable) dimensional linear programs is subtle. Strong duality and even *weak* duality may fail to hold, even for infinite linear programs with a countable number of variables and constraints [22]. Here, our primal variables live in the space of bounded measurable functions, and there are an uncountably infinite set of constraints: it is a *continuous* linear program (see [25] for a review of some of the relevant literature). Fortunately, the particular structure of our continuous linear program is of a well-behaved form.

A continuous time linear program [21] is (after a possible change of variables) of the form

$$
\min \quad \int_{\tilde{T}_0}^{\tilde{T}_1} \tilde{c}^{\top}(\theta) y(\theta) d\theta
$$

$$\begin{aligned} \text{s.t.} \quad & \tilde{B}(\theta)y(\theta) \geq \tilde{b}(\theta) + \int_{\tilde{T}_0}^{\theta} \tilde{K}(\xi,\theta)y(\xi)d\xi \qquad \forall\ \tilde{T}_0 \leq \theta \leq \tilde{T}_1 \\ & y(\theta) \geq 0 \qquad \forall\ \tilde{T}_0 \leq \theta \leq \tilde{T}_1. \end{aligned}$$

Here, $[\tilde{T}_0, \tilde{T}_1]$ is a compact interval, $\tilde{c}$ and $\tilde{b}$ are vectors of bounded measurable functions, and $\tilde{B}$, $\tilde{K}$ are matrices of bounded measurable functions. Note that $\tilde{K}(\xi,\theta)$ is only required for $\xi \leq \theta$. The components of a feasible solution y are required also to be bounded and measurable. The corresponding dual program is

$$\begin{aligned} \max \quad & \int_{\tilde{T}_0}^{\tilde{T}_1} \tilde{b}^\top(\theta)w(\theta)d\theta \\ \text{s.t.} \quad & \tilde{B}^\top(\theta)w(\theta) \leq \tilde{c}(\theta) + \int_{\theta}^{\tilde{T}_1} \tilde{K}^\top(\theta,\xi)w(\xi)d\xi \qquad \forall\ \tilde{T}_0 \leq \theta \leq \tilde{T}_1 \\ & w(\theta) \geq 0 \qquad \forall\ \tilde{T}_0 \leq \theta \leq \tilde{T}_1. \end{aligned}$$

Strong duality does not hold without further assumptions, but weak duality (and hence sufficiency of related complementary slackness conditions) do hold [21, Theorem 1]. As a consequence, if solutions y and w are feasible to the primal and dual, and satisfy

$$\begin{aligned} \int_{\tilde{T}_0}^{\tilde{T}_1} \left[w^\top(\theta)\left(\tilde{b}(\theta) - \int_{\tilde{T}_0}^{\theta} \tilde{K}(\xi,\theta)y(\xi)d\xi - \tilde{B}(\theta)y(\theta)\right)\right] = 0, \\ \int_{\tilde{T}_0}^{\tilde{T}_1} \left[y^\top(\theta)\left(\tilde{c}(\theta) - \int_{\theta}^{\tilde{T}_1} \tilde{K}(\theta,\xi)w(\xi)d\xi - \tilde{B}^\top(\theta)w(\theta)\right)\right] = 0, \end{aligned}$$

then y and w are both optimal. Reiland [21] gives constraint qualifications under which a version of this is both necessary and sufficient for optimality, i.e., where strong duality holds; we will not need this, and so don't discuss this further here.

Our continuous LP (5) fits within this class. First, we note that while we wrote the program with an unbounded interval, this was purely for notational convenience; any optimal solution must be contained in the interval $\{\theta : |\theta| \leq Q/\nu_{\text{SP}} + \tau_{\text{SP}}\}$, where ν_{SP} is the minimum capacity of an arc of some shortest s-t-path in G, and τ_{SP} is the length of this path. (This comes from considering a solution that minimizes the average journey time, and has minimum scheduling cost subject to this.) One may introduce the additional variables $F_e(\theta) = \int_{-\infty}^{\theta} f_e(\xi)d\xi$, after which it is straightforward to place things in the desired form.

After writing down the dual constraints and the complementary slackness conditions, the following sufficient conditions for optimality to (5) are obtained. The theorem is stated only for the case where the primal flow has no waiting: our algorithm produces such a flow, and so this is the case of interest to us (it will thus be an immediate corollary of our result that there is always an optimal flow without waiting). For completeness and convenience, we give a short, self-contained proof of this theorem; none of the above discussion will be used.

Theorem 4 *Let f be a flow over time without waiting and with value Q, and suppose that $\pi : V \times \mathbb{R} \to \mathbb{R}$ satisfies the following, for some choice of C:*

(i) $\theta \to \pi_v(\theta) - \alpha\theta$ *is nonincreasing.*
(ii) $\pi_w(\theta + \tau_{vw}) \leq \pi_v(\theta) + \alpha\tau_{vw}$ *for all* $\theta \in \mathbb{R}$, $vw \in E^f(\theta)$.
(iii) $\pi_s(\theta) = 0$ *for all* $\theta \in \mathbb{R}$.
(iv) $\pi_t(\theta) = (C - \rho(\theta))^+$ *for all* $\theta \in \mathbb{R}$, *and* $\nabla f_t(\theta) = 0$ *whenever* $\rho(\theta) > C$.

Then f is an optimal solution.

Proof We will need the following technical lemma (obvious via integrating by parts in the case that h is also absolutely continuous).

Claim 5 *Let $h : \mathbb{R} \to \mathbb{R}$ be a nonincreasing function, and $z : \mathbb{R} \to \mathbb{R}_+$ be an absolutely continuous nonnegative function with compact support. Then $\int_{-\infty}^{\infty} h(\theta)z'(\theta)d\theta \geq 0$.*

Proof Since z is absolutely continuous with compact support, $\int_{-\infty}^{\infty} z'(\theta)d\theta = \lim_{R\to\infty} z(R) - z(-R) = 0$. Thus the integral in the claim is invariant to replacing h with $\theta \to h(\theta) + C$ for any constant C, and so we may assume without loss of generality that h is nonnegative on the support of z. Let μ be a measure so that $\mu([\theta, \infty)) = h(\theta)$ for almost every θ in the support of z. We certainly have that for any θ,

$$\int_{-\infty}^{\theta} z'(\xi)d\xi = \left[z(\xi)\right]_{-\infty}^{\theta} = z(\theta) - 0 \geq 0.$$

Thus

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \mathbf{1}_{\xi\leq\theta} z'(\xi)d\xi d\mu(\theta) \geq 0.$$

But Fubini's theorem tells us that

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \mathbf{1}_{\xi\leq\theta} z'(\xi)d\xi d\mu(\theta) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \mathbf{1}_{\xi\leq\theta} z'(\xi)d\mu(\theta)d\xi = \int_{-\infty}^{\infty} h(\xi)d\xi,$$

which proves the claim. □

Define, for each $vw \in E$,

$$\mu_{vw}(\theta) := (\pi_w(\theta + \tau_{vw}) - \pi_v(\theta) - \alpha\tau_{vw})^+.$$

Now let g, z be any feasible solution to (5) with compact support. Consider any $v \in V \setminus \{s, t\}$, and observe that

$$\begin{aligned}\int_{-\infty}^{\infty} \pi_v(\theta)\nabla g_v(\theta) + \alpha z_v(\theta)d\theta &= \int_{-\infty}^{\infty} (\pi_v(\theta) - \alpha\theta)\nabla g_v(\theta)d\theta \;+\; \alpha\int_{-\infty}^{\infty} \theta\nabla g_v(\theta) + z_v(\theta)d\theta \\ &= \int_{-\infty}^{\infty} (\pi_v(\theta) - \alpha\theta)\nabla g_v(\theta)d\theta \;+\; \left[\alpha\theta z_v(\theta)\right]_{-\infty}^{\infty} \\ &\geq 0.\end{aligned} \tag{6}$$

The final inequality comes from observing that the first term is nonnegative by Claim 5 (applied with $h(\theta) = \pi_v(\theta) - \alpha\theta$, which is nonincreasing by property (i), and $z(\theta) = z_v(\theta)$), and that the second term is zero since z has compact support.

We then have the following sequence of inequalities (detailed explanations for each step follow).

$$\begin{aligned}
\text{cost}(g) &= \int_{-\infty}^{\infty} \rho(\theta)\nabla g_t(\theta)d\theta + \sum_{e\in E}\int_{-\infty}^{\infty} \alpha\tau_e g_e(\theta)d\theta + \alpha \sum_{v\in V\setminus\{s,t\}}\int_{-\infty}^{\infty} z_v(\theta)d\theta \\
&\overset{(*)}{\geq} \int_{-\infty}^{\infty} (C - \pi_t(\theta))\nabla g_t(\theta)d\theta + \sum_{vw\in E}\int_{-\infty}^{\infty} \big(\pi_w(\theta+\tau_{vw}) - \pi_v(\theta) - \mu_{vw}(\theta)\big)g_{vw}(\theta)d\theta \\
&\quad + \alpha \sum_{v\in V\setminus\{s,t\}}\int_{-\infty}^{\infty} z_v(\theta)d\theta \\
&\overset{(**)}{=} CQ + \sum_{v\in V\setminus\{s,t\}}\int_{-\infty}^{\infty} [\pi_v(\theta)\nabla g_v(\theta) + \alpha z_v(\theta)]d\theta - \sum_{e\in E}\int_{-\infty}^{\infty} \mu_e(\theta)g_e(\theta)d\theta \\
&\overset{(***)}{\geq} CQ - \sum_{e\in E}\int_{-\infty}^{\infty} \mu_e(\theta)\nu_e d\theta.
\end{aligned}$$

Inequality $(*)$ follows from property (iv) of π, along with the definition of μ_e. The equality $(**)$ follows by recombining the $g_e(\theta)$ terms and recalling that $\pi_s \equiv 0$ and that g has value Q. Finally, $(***)$ follows from (6), and the inequalities $\mu_e(\theta) \geq 0$ and $g_e(\theta) \leq \nu_e$ that hold for all $e \in E$ and $\theta \in \mathbb{R}$.

To complete the proof of the theorem, we now observe that all of the inequalities in the above hold with equality if $g = f$ and (consistent with the no-waiting assumption on f) $z = 0$. Property (ii) implies that if $f_{vw}(\theta) > 0$ (so that $wv \in E^f(\theta)$), then $\mu_{vw}(\theta) = \pi_w(\theta + \tau_{vw}) - \pi_v(\theta) - \alpha\tau_{vw}$, yielding equality in $(*)$. It also implies that if $f_{vw}(\theta) < \nu_{vw}$ (so that $vw \in E^f(\theta)$) then $\mu_{vw}(\theta) = 0$. This, together with $z = 0$, implies the equality in $(***)$. □

As is often the case, the optimal dual solution also provides us the prescription for tolls to induce the optimum flow. We delay this discussion to Sect. 5.

4.2 The dual prescription

We now give a certificate of optimality $\pi : V \times \mathbb{R} \to \mathbb{R}$ for (5) that satisfies the conditions of Theorem 4. Given a vertex $v \in V$ and a time $\theta \in \mathbb{R}$ let

$$\pi_v(\theta) = \max\{\hat{\pi}_v(\theta), \bar{\pi}_v(\theta), 0\}$$

where

$$\begin{aligned}
\hat{\pi}_v(\theta) &= -\alpha d_{J(v,\theta)}(v, s), \\
\bar{\pi}_v(\theta) &= C - \alpha d_{J(v,\theta)}(v, t) - \rho(\theta + d_{J(v,\theta)}(v, t)).
\end{aligned}$$

Some intuition for this choice of π can be obtained by thinking in terms of "temporal" shortest paths in the residual $E^f(\theta)$ of the flow f returned by the algorithm. For

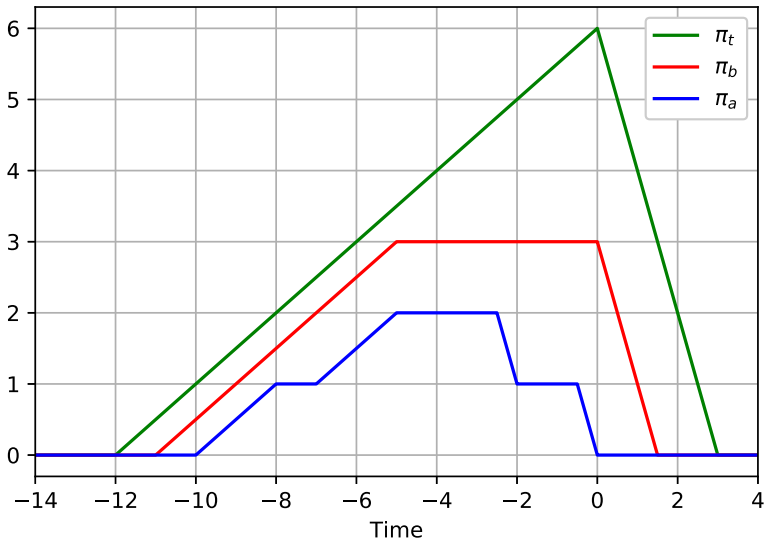


Fig. 4 Dual values to show optimality of the flow in Example 1. $\pi_s(\theta) = 0$ for all $\theta \in \mathbb{R}$, and is not shown

some $v \in V$ and $\theta \in \mathbb{R}$, consider a shortest s-v-path $P = (s = v_0, v_1, \ldots, v_{k-1}, v_k = v)$ in $f^{(J(v,\theta))}$. This path can be turned into a "temporal" path that ends at v at time θ, in the obvious way: the path should visit node v_i at time θ_i, such that $\theta_i = \theta_{i-1} + \tau_{v_{i-1}v_i}$ for each $i \in [k]$, and $\theta_k = \theta$. It turns out that every arc in this temporal path lies in the residual $E^f(\theta)$, and given that $\pi_s \equiv 0$, this implies an upper bound on $\pi_v(\theta)$ by Theorem 4 property (ii); this upper bound motivates the definition of $\hat{\pi}$. A similar consideration of a shortest t-v-path in $G^{(J(v,\theta))}$, along with the requirement that $\pi_t(\theta) = (C - \rho(\theta))^+$, motivates $\bar{\pi}$.

Example 2 Figure 4 shows the dual values π_v for the instance of Example 1. The conditions for Theorem 4 are all satisfied, for example:

- $0 < f_{sb}(-7) < \nu_{sb}$, and so we should have that $\pi_b(-7+3) = \pi_s(-7) + \alpha \cdot 3 = 3$, which is indeed the case.
- $0 < f_{sa}(\theta)$ for $\theta \in [-9, -2)$, and so we should have that $\pi_a(\theta + 1) \geq \pi_s(-\theta) + \alpha \cdot 1 = 1$ in this interval, which indeed holds.

Lemma 6 *We have $\pi_s(\theta) = 0$ and $\pi_t(\theta) = (C - \rho(\theta))^+$ for all $\theta \in \mathbb{R}$.*

Proof Notice that

$$\begin{aligned} &\hat{\pi}_s(\theta) = -\alpha d_{J(s,\theta)}(s, s) = 0 \\ \text{and} \quad &\bar{\pi}_s(\theta) = C - \alpha d_{J(s,\theta)}(s, t) - \rho(\theta + d_{J(s,\theta)}(s, t) < 0; \end{aligned}$$

the last inequality comes from (4). Thus $\pi_s(\theta) = 0$ for all θ.

Next, set $j = J(t, \theta)$ and observe that

$$\bar{\pi}_t(\theta) = C - \alpha d_j(t, t) - \rho(\theta + d_j(t, t)) = C - \rho(\theta).$$

For $\hat{\pi}_t(\theta)$, we consider two cases.

- If $j = 0$, then $\hat{\pi}_t(\theta) = -\alpha d_0(t, s) \leq 0$.
- If $j \geq 1$, then by (3) (and using $d_{j-1}(t, t) = 0$),

$$\alpha d_{j-1}(s, t) + \rho(\theta) \leq C.$$

  Since $d_{j-1}(s, t)$ is equal to the length of path P_j, and the reverse of P_j is a t-s-path in G_j, we deduce that $d_j(t, s) \leq -d_{j-1}(t, s)$, and hence that $\hat{\pi}_t(\theta) = \alpha d_j(t, s) \leq C - \rho(\theta)$.

In either case, $\pi_t(\theta) = \max\{C - \rho(\theta), 0\}$. □

Thus conditions (iii) and (iv) of Theorem 4 hold. For the remaining conditions, we begin with some basic facts about distance labels associated with successive shortest paths (statements of a similar flavour can be found in Ahuja et al. [1], for example).

Lemma 7 *For every $v \in V$, $d_j(v, s)$ is nonincreasing in j, and $d_j(v, t)$ is nondecreasing in j.*

Proof We show that $d_{j-1}(v, s) \geq d_j(v, s)$ for all $v \in V$ and $j \in [m]$. If $d_{j-1}(v, s)$ is infinite, there is nothing to prove; by possibly restricting to a subgraph in the following argument, assume that all nodes can reach s in G_{j-1}. For any node labels $\sigma \in \mathbb{R}^V$, and any $vw \in \overleftrightarrow{E}$, let $c^\sigma_{vw} := c_{vw} + \sigma_w - \sigma_v$. Notice that for any $\ell \in [m]$ and $u, z \in V$ with z reachable from u in G_ℓ, if σ is such that $c^\sigma_{vw} \geq 0$ for all $vw \in E_\ell$, then by summing the constraints along a shortest path from any u to z, $\sigma_z - \sigma_u \geq d_\ell(u, z)$. (Note that we make use of the fact that G_ℓ has no negative cost cycles.)

Now define $\sigma_v = -d_j(v, s)$ for each node v; then $c^\sigma_{vw} = c_{vw} - d_j(w, s) + d_j(v, s) \geq 0$ for all $vw \in E_j$, with equality if w lies on a shortest path from v to s. P_j is a shortest path from s to t in G_{j-1}, and hence it is a shortest path from t to s in G_j. Thus all arcs vw in the path P_j satisfy $c^\sigma_{vw} = 0$. Hence $c^\sigma_{vw} \geq 0$ for all $vw \in E_{j-1}$, implying that $d_j(v, s) = -\sigma_v \leq d_{j-1}(v, s)$ for all $v \in V$.

A similar argument "in reverse" applies for distances to t. This time, define $\sigma_v = -d_{j-1}(v, t)$; then $c^\sigma_{vw} \geq 0$ for all $vw \in E_{j-1}$, with equality for all arcs of P_j. So $c^\sigma_{vw} \geq 0$ for all $vw \in E_j$, and hence $d_{j-1}(v, t) = \sigma_t - \sigma_v \geq d_j(v, t)$ for all nodes v. □

Lemma 8 *For all $j \in [m]$ and $v \in V$, $d_{j-1}(v, t) - d_{j-1}(s, t) = d_j(v, s)$.*

Proof First of all, notice that $d_{j-1}(v, t)$ is finite precisely if $d_j(v, s)$ is, since a v-t-path in G_{j-1} can be combined with the reverse of P_j to obtain a v-s-path in G_j, and vice versa. Since $d_{j-1}(s, t)$ is always finite, the claim holds if $d_{j-1}(v, t) = d_j(v, s) = \infty$, so we assume both are finite in what follows.

We continue along the lines of the proof of the previous lemma, and define c^σ for $\sigma \in \mathbb{R}^V$ in the same way. Let $\sigma_w = -d_{j-1}(w, t)$ for all $w \in V$. Then (just as in the proof of Lemma 7) $c^\sigma_{uw} \geq 0$ for all $uw \in E_{j-1}$, and $c^\sigma_{uw} = 0$ for all $uw \in E_{j-1}$ lying on a shortest path from any node to t in G_{j-1} (including all arcs in P_j). This implies that $c^\sigma_{uw} \geq 0$ for all $uw \in E_j$, with equality for arcs in $\overleftarrow{P_j}$.

Now consider a shortest path from v to t in G_{j-1}, and let R be the segment of this path from v until the first time a node in P_j is encountered, at some node z. Let Q be the v-s-path obtained by concatenating R and the portion of $\overleftarrow{P_j}$ from z to s; Q is a path in G_j. Further, $c^{\sigma}_{uw} = 0$ for all arcs uw in Q. Thus $d_j(v,s) = \sigma_s - \sigma_v = d_{j-1}(v,t) - d_{j-1}(s,t)$, as required. □

Now we are ready to show that π satisfies conditions (i) and (ii) of Theorem 4.

Lemma 9 *$\theta \to \pi_v(\theta) - \alpha\theta$ is nonincreasing.*

Proof Fix any $\theta \in \mathbb{R}$ and $\epsilon \geq 0$. We show that $\pi_v(\theta) \geq \pi_v(\theta+\epsilon) - \alpha\epsilon$. Let $j := J(v,\theta)$ and $\ell := J(v,\theta+\epsilon)$.

- **Case 1:** $\pi_v(\theta+\epsilon) = -\alpha d_\ell(v,s)$.

  If $\ell \leq j$, then by Lemma 7

  $$\pi_v(\theta) \geq \hat{\pi}_v(\theta) = -\alpha d_j(v,s) \geq -\alpha d_\ell(v,s) = \pi_v(\theta+\epsilon).$$

  So suppose $\ell > j$. By the definition of $J(v,\theta+\epsilon)$, we know that

  $$\alpha d_{\ell-1}(s,t) + \rho(\theta+\epsilon+d_{\ell-1}(v,t)) \leq C. \tag{7}$$

  As a consequence, we have that:

  $$\begin{aligned}
  \pi_v(\theta) \geq& \bar{\pi}_v(\theta) \\
  =& C - \alpha d_j(v,t) - \rho(\theta + d_j(v,t)) \\
  \overset{(*)}{\geq}& C - \alpha d_j(v,t) - \rho(\theta+\epsilon+d_{\ell-1}(v,t)) \\
  & - \alpha\left(\epsilon + d_{\ell-1}(v,t) - d_j(v,t)\right) \\
  \geq& \alpha d_{\ell-1}(s,t) - \alpha\epsilon - \alpha d_{\ell-1}(v,t) && \text{by (7)} \\
  =& -\alpha\epsilon - \alpha d_\ell(v,s) && \text{by Lemma 8} \\
  =& \pi_v(\theta+\epsilon) - \alpha\epsilon.
  \end{aligned}$$

  Inequality $(*)$ follows from the growth bound on ρ combined with the fact that $\theta+\epsilon+d_{\ell-1}(v,t) \geq \theta + d_j(v,t)$ by Lemma 7.
- **Case 2:** $\pi_v(\theta+\epsilon) = C - \alpha d_\ell(v,t) - \rho(\theta+\epsilon+d_\ell(v,t))$.

  If $\ell \geq j$, then:

  $$\begin{aligned}
  \pi_v(\theta) &\geq \bar{\pi}_v(\theta) \\
  &= C - \alpha d_j(v,t) - \rho(\theta + d_j(v,t)) \\
  &= C - \alpha d_j(v,t) - \rho(\theta+\epsilon+d_\ell(v,t)) + \rho(\theta+\epsilon+d_\ell(v,t)) - \rho(\theta+d_j(v,t)) \\
  &\geq C - \alpha d_j(v,t) - \rho(\theta+\epsilon+d_\ell(v,t)) - \alpha\left(\epsilon + d_\ell(v,t) - d_j(v,t)\right) \\
  &= C - \rho(\theta+\epsilon+d_\ell(v,t)) - \alpha\epsilon - \alpha d_\ell(v,t)
  \end{aligned}$$

$$= \pi_v(\theta + \epsilon) - \alpha\epsilon.$$

The second inequality follows again from the growth bound, this time combined with the inequality $d_\ell(v,t) \geq d_j(v,t)$.

If $\ell < j$, by definition of $J(v, \theta + \epsilon)$ we have that

$$\alpha d_\ell(s,t) + \rho(\theta + \epsilon + d_\ell(v,t)) > C.$$

From this, we obtain

$$\begin{aligned}
\pi_v(\theta) &\geq \hat{\pi}_v(\theta) \\
&= -\alpha d_j(v,s) \\
&> C - \alpha d_\ell(s,t) - \rho(\theta + \epsilon + d_\ell(v,t)) - \alpha d_j(v,s) \\
&\geq C - \alpha d_\ell(s,t) - \rho(\theta + \epsilon + d_\ell(v,t)) - \alpha d_{\ell+1}(v,s) && \text{by Lemma 7} \\
&= C - \alpha d_\ell(v,t) - \rho(\theta + \epsilon + d_\ell(v,t)) && \text{by Lemma 8} \\
&= \pi_v(\theta + \epsilon).
\end{aligned}$$

- **Case 3:** $\pi_v(\theta + \epsilon) = 0$.

  This case is immediate from the definition of π_v. □

Lemma 10 *If $vw \in E^f(\theta)$, then $\pi_w(\theta + \tau_{vw}) \leq \pi_v(\theta) + \alpha\tau_{vw}$.*

Proof Let $j := J(v,\theta)$ and $\ell := J(w, \theta + \tau_{vw})$. Note that since $vw \in E^f(\theta)$, Theorem 1 implies that $vw \in E_j$.

- **Case 1:** $\pi_w(\theta + \tau_{vw}) = -\alpha d_\ell(w,s)$.

  If $\ell \leq j$, then

$$\begin{aligned}
\pi_v(\theta) &\geq -\alpha d_j(v,s) \\
&\geq -\alpha\tau_{vw} - \alpha d_j(w,s) && \text{since } vw \in E_j \\
&\geq -\alpha\tau_{vw} - \alpha d_\ell(w,s) && \text{by Lemma 7} \\
&= \pi_w(\theta + \tau_{vw}) - \alpha\tau_{vw}.
\end{aligned}$$

  So suppose $\ell > j$. By the definition of $J(w, \theta + \tau_{vw})$ we know that

$$\alpha d_{\ell-1}(s,t) + \rho(\theta + \tau_{vw} + d_{\ell-1}(w,t)) \leq C. \tag{8}$$

  Since $vw \in E_j$ and $d_j(w,t) \leq d_{\ell-1}(w,t)$ by Lemma 7, we also have

$$\theta + d_j(v,t) \leq \theta + \tau_{vw} + d_{\ell-1}(w,t). \tag{9}$$

Thus

$$\begin{aligned}
\pi_v(\theta) \geq & C - \alpha d_j(v,t) - \rho(\theta + d_j(v,t)) \\
\geq & C - \alpha d_j(v,t) - \rho(\theta + \tau_{vw} + d_{\ell-1}(w,t)) \\
& - \alpha \left(\tau_{vw} + d_{\ell-1}(w,t) - d_j(v,t)\right) \\
\geq & \alpha d_{\ell-1}(s,t) - \alpha\tau_{vw} - \alpha d_{\ell-1}(w,t) && \text{by (8)} \\
= & -\alpha\tau_{vw} - \alpha d_\ell(w,s) && \text{by Lemma 8} \\
= & \pi_w(\theta + \tau_{vw}) - \alpha\tau_{vw}
\end{aligned}$$

where the second inequality follows from the growth bound and from (9).

- **Case 2:** $\pi_w(\theta + \tau_{vw}) = C - \alpha d_\ell(w,t) - \rho(\theta + \tau_{vw} + d_\ell(w,t))$.

If $\ell \geq j$, since $vw \in E_j$ and $d_j(w,t) \leq d_\ell(w,t)$ by Lemma 7, we have that

$$\theta + d_j(v,t) \leq \theta + \tau_{vw} + d_\ell(w,t). \tag{10}$$

As a consequence, exploiting also the growth bound, we have

$$\begin{aligned}
\pi_v(\theta) \geq & \bar{\pi}_v(\theta) \\
= & C - \alpha d_j(v,t) - \rho(\theta + d_j(v,t)) \\
= & C - \alpha d_j(v,t) - \rho(\theta + \tau_{vw} + d_\ell(w,t)) \\
& + \rho(\theta + \tau_{vw} + d_\ell(w,t)) - \rho(\theta + d_j(v,t)) \\
\geq & C - \alpha d_j(v,t) - \rho(\theta + \tau_{vw} + d_\ell(w,t)) - \alpha\left(\tau_{vw} + d_\ell(w,t) - d_j(v,t)\right) \\
= & C - \rho(\theta + \tau_{vw} + d_\ell(w,t)) - \alpha\tau_{vw} - \alpha d_\ell(w,t) \\
= & \pi_w(\theta + \tau_{vw}) - \alpha\tau_{vw}.
\end{aligned}$$

If $\ell < j$, by definition of $J(w, \theta + \tau_{vw})$ we have that

$$\alpha d_\ell(s,t) + \rho(\theta + \tau_{vw} + d_\ell(w,t)) > C. \tag{11}$$

Thus

$$\begin{aligned}
\pi_v(\theta) \geq & -\alpha d_j(v,s) \\
\geq & -\alpha d_j(w,s) - \alpha\tau_{vw} && \text{since } vw \in E_j \\
\geq & -\alpha d_{\ell+1}(w,s) - \alpha\tau_{vw} && \text{by Lemma 7} \\
> & C - \alpha d_\ell(s,t) - \rho(\theta + \tau_{vw} + d_\ell(w,t)) - \alpha d_{\ell+1}(w,s) - \alpha\tau_{vw} && \text{by (11)} \\
= & C - \alpha d_\ell(w,t) - \rho(\theta + \tau_{vw} + d_\ell(w,t)) - \alpha\tau_{vw} && \text{by Lemma 8} \\
= & \pi_w(\theta + \tau_{vw}) - \alpha\tau_{vw}.
\end{aligned}$$

- **Case 3:** $\pi_w(\theta + \tau_{vw}) = 0$.

By Lemma 9, we have $\pi_v(\theta) + \alpha\tau_{vw} \geq \pi_v(\theta + \tau_{vw})$, which is nonnegative by the definition of π_v. □

This completes the proof that π satisfies all conditions of Theorem 4 with respect to the flow over time f produced by the algorithm, hence demonstrating the optimality of the algorithm.

5 Optimal tolls

Tolls $\mu : E \times \mathbb{R} \to \mathbb{R}_+$ are per-arc, time-varying and nonnegative. The value $\mu_e(\xi)$ represents the toll a user is charged upon entering the link at time ξ.

We have the following theorem.

Theorem 11 *Let (f, π) be an optimal primal-dual solution to (5) (as constructed in Sects. 3 and 4) and define, for each $vw \in E$,*

$$\mu_{vw}(\theta) = (\pi_w(\theta + \tau_{vw}) - \pi_v(\theta) - \alpha\tau_{vw})^+.$$

Then f is a dynamic equilibrium under tolls μ.

Of course, to make sense of this theorem we must know what is meant by a dynamic equilibrium under tolls. Informally, it means that no user (represented as an infinitesimal flow particle) has an alternative strategy (route choice and departure time combination) of strictly smaller disutility. Making this precise in general requires defining precisely what disutility a user would incur for any given route and departure time choice, by considering the full game-theoretic fluid queueing model (also known as the Vickrey bottleneck model) [16,28]. Tolls and departure time choice can be introduced into the definition of a dynamic equilibrium discussed in these works. However, the complexity of this is only needed because in general, a user may have to incur additional waiting time on arcs that are fully utilized, meaning that the disutility of a particular strategy depends in a complicated way on the actions of the other users. Instead we show that *even* if a user is allowed to traverse any link at any time—as if the other users were not present—there is no incentive to deviate. This is clearly a stronger property than any reasonable notion of equilibrium.

Let C be the cost horizon associated with (f, π). We will show that all users experience a disutility of exactly C with f under tolls μ, and in addition that the disutility of any other possible choice is at least C. To see this, consider any s-t-path P in G, along with departure times θ_v for each $v \in P$ valid for this path, meaning that $\theta_w = \epsilon + \theta_v + \tau_{vw}$ for every $vw \in P$, with $\epsilon \geq 0$. (So we allow the possibility of waiting at a vertex). Thus by the definition of μ and by properties (i) and (ii) of Theorem 4,

$$\mu_{vw}(\theta_v + \epsilon) + \alpha(\tau_{vw} + \epsilon) \geq \pi_w(\theta_w) - \pi_v(\theta_v + \epsilon) + \alpha\epsilon \geq \pi_w(\theta_w) - \pi_v(\theta_v)\ ,$$

with equality if $wv \in E^f(\theta_w)$ and $\epsilon = 0$. Then the disutility of a user using this route is

$$\begin{aligned}
&\rho(\theta_t) + \sum_{e=vw \in P} [\alpha(\theta_w - \theta_v) + \mu_e(\theta_w - \tau_e)] \\
&\geq \rho(\theta_t) + \pi_t(\theta_t) - \pi_s(\theta_s) \\
&= \rho(\theta_t) + (C - \rho(\theta_t))^+ \\
&\geq C.
\end{aligned}$$

The inequalities are all tight if for all $vw \in P, \theta_w = \theta_v + \tau_{vw}$ and $f_{vw}(\theta_v) > 0$, by the previous observations as well as property (iv). So if the aggregate choices of the users are described by f, all users pay exactly C. This completes the proof of Theorem 11.

5.1 Strong vs weak enforcement

As already discussed, we cannot in general strongly enforce an optimal flow, i.e., set tolls such that every dynamic equilibrium is optimal. The following shows that the "lane tolling" approach suffices to do this.

Theorem 12 *Let f, π and μ be as in the previous theorem, and suppose g is any dynamic equilibrium satisfying $g_e(\theta) \leq f_e(\theta)$ for all $e \in E$, $\theta \in \mathbb{R}$. Then g is optimal.*

Proof The cost of g cannot exceed the cost of f, and so it must be optimal. □

Essentially, being able to dynamically split and separately toll the capacity of a link allows us to easily rule out all other potential equilibria just by using tolls to artificially constrict the capacities (in addition to choosing tolls that weakly enforce the desired flow, which is still needed). Tolling in this way seems quite distant from what could be imaginable in realistic traffic scenarios. But it does raise the interesting question of whether there is a tolling scheme which can strongly enforce an optimum flow, but which is more restricted (and more plausible) than fully dynamic lane tolling. Another natural question would be to determine if an optimum flow can be strongly enforced using lane tolling only on certain specified edges. We leave these as open questions.

5.2 Exogenous demand

Now let us consider the case of exogenous, and fixed, demand. Users depart from the source s at a fixed rate u_0 over a time interval $[0, T]$, and simply wish to reach the destination t as early as possible. Correspondingly, the social cost we wish to minimize is the average journey time. Note that there is no longer any departure time choice; as such, users departing at different times need not experience the same disutility (journey time) in an equilibrium.

We can view this within our setting as follows. Let $G' = (V', E')$ be the instance obtained by adding a node s' and an arc $s's$ of capacity u_0 and delay 0; s' becomes the new source. The total flow to send is $Q = Tu_0$. The arc $s's$ ensures that the amount

of flow departing s by time θ in any flow over time cannot exceed $u_0\theta$. A flow over time need not saturate the arc $s's$ in the interval $[0, T]$; however, as long as a given flow is nonzero only for nonnegative times, we can easily convert it to one that does. Simply adjust the flow to send flow from s' to s earlier, saturating $s's$ on $[0, T]$, and then waiting at s (that is, we do not modify the flow on any other arcs). This clearly has no impact on arrival times. As such, we can view the restriction to G of any flow over time on G' (that is nonzero only for nonnegative times) as a potential solution to the exogenous demand problem.

Now let $\bar{G}$ be obtained from G' by reversing all arcs. We consider now the source to be t and the sink s', and the scheduling cost function described in (2). Let $\bar{f}$ be an optimal flow of value $Q = Tu_0$, and let $\bar{\mu}$ be the optimal tolls from Theorem 11 that induce it.

We now "reverse time". For any arc $vw \in E'$, and $\theta \in \mathbb{R}$, let $f_{vw}(\theta) = \bar{f}_{wv}(\tau_{vw} - \theta)$, and let $\mu_{vw}(\theta) = \bar{\mu}_{wv}(\tau_{vw} - \theta)$. Then f is an earliest arrival flow in G': f is zero for all negative times (since all flow in $\bar{f}$ arrives by time 0), and since $\bar{f}$ minimizes the average departure time from t in $\bar{G}$, f minimizes the average arrival time at t in G'.

Lemma 13 *The tolls μ restricted to E induce the restriction of f to G as a dynamic equilibrium under exogenous demand.*

Proof First, we observe that the tolls μ induce f as a dynamic equilibrium in G'. This is simply because given any s'-t-path P and valid departure times $(\theta_v)_{v\in P}$, these can be mapped to a reversed path $\bar{P}$ from t to s' in $\bar{G}$, along with corresponding reversed departure times $(\bar{\theta}_v)_{v\in\bar{P}}$ (given by, for any $vw \in P$, $\bar{\theta}_w = \tau_{vw} - \theta_v$). The disutility experienced by a user in G' choosing the strategy described by P and $(\theta_v)_{v\in P}$ is then precisely equal to the disutility experienced by a user in $\bar{G}$ choosing the strategy described by $\bar{P}$ and $(\bar{\theta}_v)_{v\in P}$.

All that remains is to go from G' to G; that is, we need to argue that the restriction of μ to V does induce the restriction of f to G. The role of $\mu_{s's}$ in G' is only to ensure equal costs between particles that traverse $s's$ at different times. This is not a requirement of an equilibrium in the exogenous setting. Some care is required however, since by restricting f to G, we are possibly introducing waiting at s.

Let $\theta' = \inf\{\theta \geq 0 : f_{s's}(\theta) < u_0\}$. Then $\mu_{s's}(\theta) = 0$ for all $\theta > \theta'$ with $f_{s's}(\theta) > 0$. To see this, consider any $\tilde{\theta} \in (\theta, \theta')$ for which $f_{s's}(\tilde{\theta}) < u_0$. Then $s's \in E^f(\tilde{\theta})$, implying by property (ii) of the dual solution π (see Theorem 4) that $\mu_{s's}(\tilde{\theta}) = 0$. But then if $\mu_{s's}(\theta')$ were larger than 0, it would be an improving deviation to traverse $s's$ at time $\tilde{\theta}$ and then wait at s, so this is not possible.

It follows that there is no waiting at s for users departing before time θ' in G (since f had no waiting, and $f_{s's}(\theta) = u_0$ until time θ'), whereas *all* users departing after time θ' experience the same disutility. Thus, no user has an incentive to deviate, and we have a dynamic equilibrium in G. □

6 General scheduling costs

We now consider general scheduling costs, satisfying only the growth bound assumption as well as the following fairly unrestrictive condition. We will assume that for any C, $\{\theta \in \mathbb{R} : \rho(\theta) \leq C\}$ consists of a finite number of compact intervals, and this number is uniformly bounded by some value K. Insisting that this set has finite measure ensures that the total mass associated with any given choice of cost horizon is finite. The assumption that this set is always closed, or in other words, that ρ is lower semicontinuous, is a matter of convenience, and was already assumed in the strongly unimodal case. Given a scheduling function that does not satisfy this, but with a finite number of discontinuities, the property can be obtained by adjusting ρ only at points of discontinuity, and without affecting the optimal solution. Finally, the assumption that the number of intervals is bounded by some K ensures that the algorithm has a finite description, and also rules out various pathological choices of ρ.

In order to actually implement the algorithm, oracle access to ρ will not suffice. Instead, we assume that given C, we are able to obtain the sets $\rho^{-1}((-\infty, C])$ and $\rho^{-1}(\{C\})$, described as collections of intervals. Note that $\rho^{-1}(\{C\})$ consists of a union of at most $2K$ intervals, since $\rho^{-1}(\{C\}) = \rho^{-1}((-\infty, C]) \setminus \bigcup_{\epsilon>0} \rho^{-1}((-\infty, C-\epsilon])$.

There are essentially two separate complications that arise compared to the strongly unimodal case. The first complication is that the set of arrival times where the scheduling cost is bounded by some value need no longer be an interval. The second is that the total mass corresponding to a given cost horizon C need no longer depend continuously on C, meaning that once we have found the "correct" choice of cost horizon, the algorithm as stated might send too much mass.

Let us begin by dealing with the first complication alone. So suppose that in addition to the stated restrictions, $\mu\big(\rho^{-1}((-\infty, C])\big)$ is a continuous function of C, where throughout this section $\mu(A)$ denotes the Lebesgue measure of a set A. The use of bisection search (or, if ρ is given as a piecewise constant function, perhaps parametrized search) to determine the correct cost horizon will thus not be affected. We need only describe how the algorithm and analysis for finding an optimal solution for a given cost horizon should be modified.

The principle of the algorithm remains identical to its description in Sect. 3. All that changes is that for a path P_j obtained from successive shortest paths, the set of times respecting the cost horizon C is no longer an interval. Thus, we let $I_j \subseteq \mathbb{R}$ be a maximal set such that

$$\rho(\xi + d_{j-1}(s,t)) \leq C - \alpha d_{j-1}(s,t) \quad \text{for all} \quad \xi \in I_j. \tag{12}$$

Given our assumptions on ρ, I_j is a finite set of compact intervals. The resulting flow f has precisely the same definition as before, namely $f_{vw}(\theta) = f_{vw}^{(J(v,\theta))}$, where the definition of $J(v,\theta)$ also remains unchanged. The proof of feasibility of this flow, and the proof of its optimality, both did not depend on any way on I_j being an interval, and the existing proofs stand as written. Note that the value of the flow is $\sum_{j=1}^{m} \mu(I_j)$; as expected, our current assumptions ensure that this is a continuous function of C.

Let us now relax the condition on continuity of $C \to \mu\big(\rho^{-1}((-\infty, C])\big)$. The essential idea is as follows.

- The algorithm as described will find the *maximum* mass corresponding to a given cost horizon C. We can also find the *minimum* corresponding mass, by slightly adjusting the algorithm.
- Once we have determined the correct value of C via bisection or parametric search, we must choose a solution that in a sense interpolates between the minimum and maximum mass solutions. Some care is required to ensure that we have a feasible flow. In particular, consider removing flow sent along a generalized path P_j coming from the SSP decomposition for some interval of time. If an edge e is a forward edge in P_j, then removing flow on it for a certain period of time may not be possible without also removing flow from a generalized path that uses e in the opposite direction.

We now describe the algorithm. As previously, a bisection search (or possibly parametrized search) is used to find the correct cost horizon. However, given a current guess C, we will compute a corresponding interval $[Q^{\min}(C), Q^{\max}(C)]$ of possible total masses corresponding to this. To do this, we construct, for each path P_j obtained from successive shortest paths, two sets $I_j^{\min}$ and $I_j^{\max}$ defined as follows. $I_j^{\max}$ is defined precisely as before, i.e., according to (12). $I_j^{\min}$ is defined instead as

$$I_j^{\min} := \mathrm{cl}\Big(I_j^{\max} \setminus \{\xi : \rho(\xi + d_{j-1}(s,t)) = C - \alpha d_{j-1}(s,t)\}\Big),$$

where $\mathrm{cl}(A)$ denotes the closure of the set $A \subseteq \mathbb{R}$. Again, our assumptions on ρ ensure that $I_j^{\min}$ is a finite collection of compact intervals. This results in two different flows over time, $f^{\min}$ and $f^{\max}$, of values

$$Q^{\min}(C) = \sum_{j=1}^{m} \mu(I_j^{\min}) \qquad \text{and} \qquad Q^{\max}(C) = \sum_{j=1}^{m} \mu(I_j^{\max})$$

respectively. (Feasibility and optimality of $f^{\min}$ has not yet been demonstrated; we will return to this point.)

Once we have found C such that $Q \in [Q^{\min}(C), Q^{\max}(C)]$, we proceed as follows. Let $\theta_0 \in \mathbb{R}$ be a value we will choose later. For each $j \in [m]$, let

$$I_j^{\theta_0} := I_j^{\min} \cup \big(I_j^{\max} \cap (-\infty, \theta_0 - d_{j-1}(s,t)]\big);$$

so $I_j^{\min} \subseteq I_j^{\theta_0} \subseteq I_j^{\max}$ (see Fig. 5 for an example). Now take f^{θ_0} to be the flow over time obtained by sending flow on path P_j for times in $I_j^{\theta_0}$, for each j. Delaying concerns about feasibility, the value of this flow is $Q(\theta_0) := \sum_{j=1}^{m} \mu(I_j^{\theta_0})$. Since this is continuous and piecewise linear in θ_0, with

$$Q^{\min}(C) = \inf_{\theta} Q(\theta) \quad \le \quad Q \quad \le \quad \sup_{\theta} Q(\theta) = Q^{\max}(C),$$

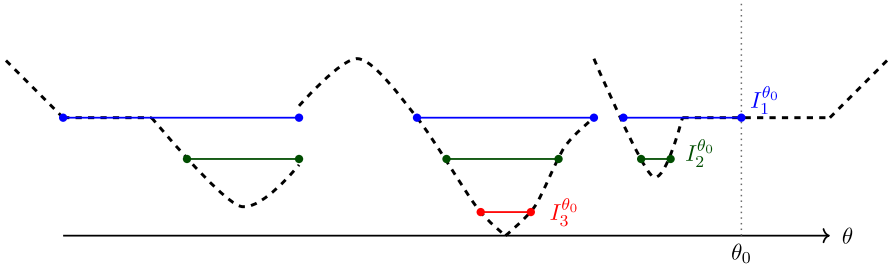


Fig. 5 An example of a general scheduling cost function, and intervals $I_j^{\theta_0}$ corresponding to three paths of different lengths, for some particular choice of C and the indicated value of θ_0

we can easily determine the correct choice for θ_0 so that $Q(\theta_0) = Q$. The output of the algorithm is then $f := f^{\theta_0}$ for this choice of θ_0.

It remains to show the correctness of this algorithm, by demonstrating that the flow over time f constructed by this algorithm is both feasible and optimal. This could be done by suitably tweaking the arguments in Sects. 3 and 4. We will however avoid this, and instead proceed as follows. Suppose we are able to define a family of perturbed scheduling cost functions $\rho^{(\epsilon)}$ and flows $f^{(\epsilon)}$ for all $\epsilon > 0$ such that the following hold:

(i) $\rho^{(\epsilon)}$ converges uniformly to ρ as $\epsilon \to 0$. Note that this implies that the cost of an optimal solution under scheduling cost $\rho^{(\epsilon)}$ converges to the cost of an optimal solution under ρ.
(ii) $f^{(\epsilon)}$ is an optimal flow of *maximum* total mass with cost horizon C, with respect to the cost function $\rho^{(\epsilon)}$, and $f^{(\epsilon)}$ converges to f as $\epsilon \to 0$.

Together, this implies feasibility and optimality of f for ρ.

We define $\rho^{(\epsilon)}$ as follows:

$$\rho^{(\epsilon)}(\theta) = \begin{cases} \rho(\theta) & \text{if } \theta \leq \theta_0 \\ \rho(\theta) + \epsilon & \text{if } \theta > \theta_0 \\ \min\{\rho(\theta_0), \lim_{\theta' \to \theta_0^+} \rho(\theta') + \epsilon\} & \text{if } \theta = \theta_0 \end{cases}.$$

(That is, we increase ρ by ϵ to the right of θ_0, choosing the value at θ_0 so that $\rho^{(\epsilon)}$ is lower semicontinuous.) Property (i) is then immediate. Property (ii) follows from (12) applied to $\rho^{(\epsilon)}$: the maximal interval corresponding to path P_j in $\rho^{(\epsilon)}$ converges to $I_j^{\theta_0}$.

Acknowledgements We thank the anonymous referees for their detailed and constructive feedback.

References

1. Ahuja, R.K., Magnanti, T.L., Orlin, J.B.: Network Flows: Theory, Algorithms, and Applications. Prentice Hall (1993)
2. Arnott, R., de Palma, A., Lindsey, R.: Economics of a bottleneck. J. Urban Econ. **27**(1), 111–130 (1990)
3. Baumann, N., Skutella, M.: Solving evacuation problems efficiently—earliest arrival flows with multiple sources. In: Proceedings of the 47th Annual IEEE Symposium on Foundations of Computer Science (FOCS), pp. 399–410 (2006)
4. Bhaskar, U., Fleischer, L., Anshelevich, E.: A Stackelberg strategy for routing flow over time. In: Proceedings of the Twenty-Second Annual ACM-SIAM Symposium on Discrete Algorithms, (SODA), pp. 192–201 (2011)
5. Cominetti, R., Correa, J., Larré, O.: Dynamic equilibria in fluid queueing networks. Oper. Res. **63**(1), 21–34 (2015)
6. Cominetti, R., Correa, J.R., Olver, N.: Long term behavior of dynamic equilibria in fluid queuing networks. In: Proceedings of the 19th International Conference on Integer Programming and Combinatorial Optimization (IPCO), pp. 161–172 (2017)
7. Correa, J.R., Cristi, A., Oosterwijk, T.: On the price of anarchy for flows over time. In: Proceedings of the 2019 ACM Conference on Economics and Computation, (EC), pp. 559–577 (2019)
8. Disser, Y., Skutella, M.: The simplex algorithm is NP-mighty. In: Proceedings of the Twenty-Sixth Annual ACM-SIAM Symposium on Discrete Algorithms, (SODA), pp. 858–872 (2015)
9. Fleischer, L., Tardos, E.: Efficient continuous-time dynamic network flow algorithms. Oper. Res. Lett. **23**(3), 71–80 (1998)
10. Ford, L.R., Fulkerson, D.R.: Constructing maximal dynamic flows from static flows. Oper. Res. **6**(3), 419–433 (1958)
11. Ford, L.R., Fulkerson, D.R.: Flows in Networks. Princeton University Press (1962)
12. Frascaria, D., Olver, N., Verhoef, E.T.: Emergent hypercongestion in Vickrey bottleneck networks. Accepted to Transportation Research Part B: Methodological. Preprint available as Tinbergen Institute Discussion Paper TI 2020-002/VIII (2020)
13. Gale, D.: Transient flows in networks. Michigan Math. J. **6**(1), 59–63 (1959)
14. Harks, T.: Pricing in resource allocation games based on lagrangean duality and convexification. Preprint, arXiv:1907.01976 (2019)
15. Jarvis, J.J., Ratliff, H.D.: Some equivalent objectives for dynamic network flow problems. Manag. Sci. **28**(1), 106–109 (1982)
16. Koch, R., Skutella, M.: Nash equilibria and the price of anarchy for flows over time. Theory Comput. Syst. **49**(1), 71–97 (2011)
17. Köhler, E., Möhring, R.H., Skutella, M.: Traffic networks and flows over time. In: Algorithmics of Large and Complex Networks - Design, Analysis, and Simulation, pp. 166–196 (2009)
18. Minieka, E.: Maximal, lexicographic, and dynamic network flows. Oper. Res. **21**(2), 517–527 (1973)
19. Nisan, N., Roughgarden, T., Tardos, É., Vazirani, V.V. (eds) Algorithmic Game Theory. Cambridge University Press (2007)
20. Philpott, A.B.: Continuous-time flows in networks. Math. Oper. Res. **15**(4), 640–661 (1990)
21. Reiland, T.W.: Optimality conditions and duality in continuous programming ii. the linear problem revisited. J. Math. Anal. Appl., **77**(2), 329–343 (1980)
22. Romeijn, H.E., Smith, R.L., Bean, J.C.: Duality in infinite dimensional linear programming. Math. Program. **53**, 79–97 (1992)
23. Sering, L., Skutella, M.: Multi-source multi-sink Nash flows over time. In: 18th Workshop on Algorithmic Approaches for Transportation Modelling, Optimization, and Systems, (ATMOS), pp. 12:1–12:20 (2018)
24. Sering, L., Koch, L.V.: Nash flows over time with spillback. In: Proceedings of the Thirtieth Annual ACM-SIAM Symposium on Discrete Algorithms, (SODA), pp. 935–945 (2019)
25. Sharkey, T.C.: Infinite Linear Programs. Wiley (2011)

26. Skutella, M.: An introduction to network flows over time. In: Research Trends in Combinatorial Optimization, pp. 451–482 (2009)
27. Small, K.A.: The bottleneck model: an assessment and interpretation. Econ. Transp. **4**(1), 110–117 (2015)
28. Vickrey, W.: Congestion theory and transport investment. Am. Econ. Rev. **59**(2), 251–60 (1969)
29. Wilkinson, W.L.: An algorithm for universal maximal dynamic flows in a network. Oper. Res. **19**(7), 1602–1612 (1971)
30. Yang, H., Meng, Q.: Departure time, route choice and congestion toll in a queuing network with elastic demand. Transp. Res. Part B: Methodol. **32**(4), 247–260 (1998)
31. Zadeh, N.: A bad network problem for the simplex method and other minimum cost flow algorithms. Math. Program. **5**, 255–266 (1973)