The line planning routing game

Philine Gattermann\textsuperscript{1}, Alexander Schiewe\textsuperscript{1}, and Marie Schmidt\textsuperscript{2}

\textsuperscript{1}: Institute for Numerical and Applied Mathematics, University of Göttingen
\textsuperscript{2}: Rotterdam School of Management, Erasmus University
Burgemeester Oudlaan 50, 3062 PA Rotterdam, The Netherlands,
phone: +31 10 4082199, email: schmidt2@rsm.nl

3rd December 2014

Abstract
In this paper, we propose a novel algorithmic approach to solve line planning problems. To
this end, we model the line planning problem as a game where the passengers are players which
aim at minimizing individual objective functions composed of travel time, transfer penalties,
and a share of the overall cost of the solution. To find equilibria of this routing game, we use
a best-response algorithm. We investigate, under which conditions on the line planning model
a passenger’s best-response can be calculated efficiently and which properties are needed to
guarantee convergence of the best-response algorithm. Furthermore, we determine the price
of anarchy which bounds the objective value of an equilibrium with respect to a system-
optimal solution of the line planning problem. For problems where best-responses cannot be
found efficiently, we propose heuristic methods. We demonstrate our findings on some small
computational examples.

Keywords: Transportation, Game Theory, Routing, Line Planning, Routing Game

1 Introduction
Due to the high complexity of public transportation planning, the planning process is normally
subdivided in subsequent steps, such as network design, line planning, timetabling, vehicle schedul-
ing, etc. The line planning problem aims at determining the routes which are served regularly by
a vehicle and the frequencies of these services. When evaluating a line concept, both the costs
for implementing the line concept and the quality of the line concept from the passengers’ perspective
have to be taken into account. Possible ways to measure the quality of a line concept $L$ from the
point of view of a passenger $p$ are the travel time and the number of transfers on the route that
passenger $p$ would choose given the line concept $L$. Hence, to evaluate a line concept $L$ from a
quality perspective, passengers’ routes for the chosen line concept have to be determined.

For this reason, recent publications on passenger-oriented line planning solve the problem of finding
a line concept $L$ and a set of passengers’ routes $R$ simultaneously, by modeling and solving it as
linear integer programs, see e.g., [SS06a, BGP07, BN10, BK12, Sch14].

In order to achieve problem formulations which can be solved for practical instances, these models
use several simplifications. Often, transfer times are assumed to be independent of line frequencies
(see, e.g., [SS06a, BN10, BK12, Sch14]) or not taken into account at all [BGP07]. [GvHK06,
GvHK04] use a model that allows to adjust transfer times to frequencies, but make a different
restriction: for each passenger, the path is the network on which he travels is fixed beforehand
(even if the exact connection, i.e., the sequence of trains used on this path, is not).

Furthermore, the cited approaches determine a system-optimum with respect of the cumulated ob-
jective functions of all passengers. In order to achieve a system-optimal solution, single passengers
may be assigned routes which are significantly worse than their individually optimal route.
One possible approach to overcome these shortcomings is to iterate between planning and routing steps, as proposed in [Sch14]. However, this heuristic allows only to remove lines from the initial solutions, but not to include new lines, so it can get stuck fast in a local optimum.

In this paper, we propose a novel algorithmic approach to solve line planning problems. Instead of integrating planning and routing steps or iterating between both as done in the approaches described above, we regard only the routing step and include all planning decisions in this step. To this end, we define an individual objective function for each passenger which is composed of travel time, transfer penalties, and a share of the overall cost of the solution. This way, the line planning problem can be interpreted as a game in which the passengers are the players who aim at minimizing their objective functions.

In this paper, we study this so-defined line planning routing game. To find equilibria we propose a best-response algorithm. We investigate, under which conditions on the line planning model a passenger’s best-response can be calculated efficiently and which properties are needed to guarantee convergence of the best-response algorithm. Furthermore, we determine the price of anarchy which bounds the objective value of an equilibrium with respect to a system-optimal solution of the line planning problem. We demonstrate our results on some numerical examples.

The remainder of this paper is structured as follows. In Section 2 we review literature on line planning before we introduce a general line planning problem (LPQC) in Sections 3.1 and 3.2. We briefly introduce some concepts from game theory in Section 3.3. In Section 3.4 we define the line planning routing game LPRG defined by LPQC and discuss the relations between the two problems in Section 3.5.

In Section 4 we investigate properties of the line planning routing game. In Section 4.1 we sketch the best-response algorithm used to find equilibria to LPRG. In Section 4.2 we investigate, under which conditions on the line planning model a passenger’s best-response can be calculated efficiently. The existence of equilibria and the convergence of the best-response algorithm is investigated in Section 4.3. Section 4.4 evaluates the solutions found by the best-response algorithm with respect to a system-optimal solution. Finally, in Section 5 we illustrate and compare the different models on some small line planning instances.

2 Related literature

Line planning is an important step in the public transportation planning process. There are many line planning models, which differ with respect to the decisions covered by the term line planning, the level of detail with which real-world constraints are included in the model, and the way of measuring the quality of a line plan. In this paper, we only give a brief overview on the line planning models and solution methods which are most relevant for this paper. See, e.g., [Sch11, Sch14] for more extensive overviews on line planning.

Line planning aims at finding a line concept (consisting of line routes and frequencies) which is good from an operational point of view and offers good quality for the passengers. Cost-oriented line planning models focus on minimizing the operational costs subject to the constraint that passenger demand has to be satisfied (see, e.g., [CvDZ98, Bus98, GvHK06, BHK+13]). Many passenger-oriented line planning models aim at minimizing the overall travel time while keeping the costs below a predefined threshold [SS06a, Sch14]. There are also some passenger-oriented models which measure quality by the number of direct travelers [Die78, Bus98, BKZ97]. Some models combine quality and cost into one objective [BGP08, GYW06, PB06].

In most literature on line planning, line planning problems are modeled and solved as integer programs. Solution approaches for cost-oriented models often assign the demand to the network edges in a preprocessing step and formulate covering or packing models. Solution techniques include branch-and-bound [Bus98, CvDZ98], branch-and-cut [GvHK04], and variable fixing heuristics [BLL04].
Passenger-oriented line planning assumes that passengers choose the "best" route with respect to the chosen line concept (where "best" is often understood as travel-time minimal). For this purpose, passengers' routes cannot be determined in a preprocessing step but have to be determined together with the line concept. [SS06a] model passengers as flows in a change-and-go network, which allows to include transfer times in the travel time, and solve the LP-relaxation using Dantzig-Wolfe decomposition. However, this leads to very large IP models and relatively long solution times. [BGP07, BN10, BK12] use column generation to generate passengers' routes. In [BN10] it is shown that this can lead to a significant speed-up with respect to flow formulations in change-and-go networks. However, transfer times are not taken into account [BGP07] or only approximated [BN10, BK12] in these models.

A drawback of the described passenger-oriented models is that they can only find system-optimal solutions, that is solutions with a minimal overall travel time. I.e., a solution to these models, consisting of line concept and passenger routes could actually require that some passengers take long detours to avoid capacity violations and allow a minimal overall travel time. [Sch14] introduces a model where only line concepts which allow all passengers to travel on shortest paths (with respect to the line concept) are considered feasible and proposes an IP formulation. However, since the formulation is based on a flow formulation in a change-and-go network, fast solution times for real-world instances are not to be expected.

Solution approaches to line planning which are not IP-based, often concentrate on the line routes only and postpone frequency setting to a later step. They use greedy strategies [CW91, PRR95, Qua03] to construct lines or successively remove lines from a big line pool [Pat25, Son79]. Furthermore, metaheuristics like genetic algorithms [FGP02, FM06a, SW11], neighborhood search [SW11], and simulated annealing [FM06b] are used. [Man80, JBT10, Sch14] describe iterative approaches, where line planning/frequency setting and route assignment steps are iterated. There are also game-theoretic approaches to line planning which model line operators as players who compete for a good utilization of the lines they offer [SS06b, Sch09, BZK11, SS13, Neu14]. In [LMP10], the problem of finding a line concept which is robust against link failures is modeled as a game between the network provider and an adversary. However, to the extent of our knowledge, so far no attempt has been made to model line planning as a game with passengers as players.

In the field of transit assignment, models from game theory are used to model passenger flows on networks (see, e.g. [SF89, She85, DCF93, NP88, SFS+11, SSJ11, CF95]). These models take into account different modeling requirements from practice, like e.g., limited seat capacity or uncertain information about the next arriving trains. Equilibria are often found by mathematical programming.

Routing games on networks are also studied from a more theoretical perspective in the area of algorithmic game theory. A good overview of this line of research, both for atomic and non-atomic flow, is given, e.g., in [Rou07]. Questions of interest cover the existence and quality of equilibria and algorithmic approaches to identify equilibria (see, e.g. [Ros73, AAE05, Rou05, ADK+04, Rou07, TW07]).

3 The line planning routing game

3.1 Line planning

Line planning aims at determining routes and frequencies of trains. As a basis, we consider the underlying public transportation network (PTN) \( G = (V,E) \). The nodes \( V \) of this network represent stations. Two stations are connected by an edge \( e \) if there is a direct track connection between the corresponding stations. In this paper, we assume that we are given a line pool \( \mathcal{L} \) of possible lines. The main task of line planning is to find a line concept, i.e., to assign a frequency \( f_l \in \mathbb{N}_0 \) to every line \( l \) in the line pool \( \mathcal{L} \).

Passenger demand is given in form of origin-destination (OD)-pairs \((u_q, v_q)\), specifying a passengers' origin \( u_q \) and his destination \( v_q \). Apart from the line concept, passenger-oriented line
planning models determine a set of passenger routes \( \mathcal{R} := \{ R_q : q \in \mathcal{Q} \} \). A route \( R_q \) for passenger \( q \) specifies a path \( P'_q = (e_1, \ldots, e_n) \) from \( u_q \) to \( v_q \) and for every \( e_i \in P'_q \) a line \( l_i \) which is used while traveling on \( e_i \). I.e., \( R_q \) can be written as a sequence \( R_q = ((e_1, l_1), (e_2, l_2), \ldots, (e_n, l_n)) \).

For a given set of routes \( \mathcal{R} \) we denote the number of passengers who use line \( l \in \mathcal{L} \) on edge \( e \) by

\[
x_{(e,l)}(\mathcal{R}) := |\{ q \in \mathcal{Q} : (e, l) \in R_q \}|
\]

A pair of frequencies \( f \) and passenger route set \( \mathcal{R} \) is feasible if for every \( l \) and every \( e \in l \) the number of passengers traveling on the line does not exceed the capacity of a single train multiplied with the frequency of the line, that is, if

\[
x_{(e,l)}(\mathcal{R}) \leq f_l B.
\]

The objective of line planning is to minimize the costs of the implemented line concept as well as its quality, estimated, e.g., by the travel time and the number of transfers. I.e., the line planning problem can be summarized as follows: given a PTN \( G \), a line pool \( \mathcal{L} \), a capacity bound \( B \), and a set of passengers \( \mathcal{Q} \): find a pair of frequencies \( f \) and routes \( \mathcal{R} \) such that

\[
x_{(e,l)}(\mathcal{R}) \leq f_l B
\]

and a weighted sum of quality and cost is minimized, as detailed in the following section.

### 3.2 The line planning objective function

There are many approaches on how to measure cost and quality, as described in the literature overview in Section 2. The models used here are described in more detail in the following:

In our model the costs \( \text{cost}_l \) of a line \( l \) are composed of a constant cost \( k^l_1 \) and a per-frequency cost \( k^l_2 \), i.e.,

\[
\text{cost}_l(f) = \begin{cases} 
  \gamma_1 \cdot k^l_1 + \gamma_2 \cdot k^l_2 \cdot f_l & \text{if } f_l > 0 \\
  0 & \text{otherwise}
\end{cases}
\]

where \( \gamma_1 \) and \( \gamma_2 \) are non-negative constants. The cost of a line concept represented by frequencies \( f \) is hence given as

\[
\text{cost}(f) := \sum_{l: f_l > 0} \text{cost}_l(f).
\]  

(1)

Quality, as perceived by a passenger, can be measured by driving time and transfer time of the passenger’s journey. We denote travel\(_q(f, R_q) := \alpha_1 c_q(R_q) + \alpha_2 \cdot \tau_q(f, R_q) \), where \( \alpha_1 \) and \( \alpha_2 \) denote non-negative weighting factors.

The driving time on route \( R_q \) depends only on the chosen route in the PTN. It is given as

\[
c_q(R_q) := \sum_{(e,l) \in R_q} c(e).
\]

The transfer time \( \tau_q(f, R_q) \) is estimated based on the frequencies of the lines involved in the transfers on the route. In this paper, for a transfer from line \( l \) to line \( l' \) we assume a transfer time of \( \frac{1}{f_l + f_{l'}} \). Then, the overall transfer time of passenger \( q \) on route \( R_q \) is

\[
\tau_q(f, R_q) := \sum_{i=1}^{n-1} \frac{1}{f_{l_i} + f_{l_{i+1}}},
\]

(2)

where \( (l_1, l_2, \ldots, l_n) \) is the sequence of lines used on \( R_q \).

Additionally, in transfer\(_q(R_q) \) a penalty \( \beta \) can be added for every transfer on \( R_q \), modeling the inconvenience arising for the passenger when transferring.
We obtain the line planning objective
\[ G(\mathcal{R}, f) := \sum_{q \in \mathcal{Q}} (\alpha_1 c_q(R_q) + \alpha_2 r_q(f, R_q)) + \sum_{q \in \mathcal{Q}} \text{transfer}_q(R_q) + \sum_{l \in \mathcal{L}, j > 0} (\gamma_1 \cdot k_1^l + \gamma_2 \cdot k_2^l f_l). \] (3)

We call the problem described in Sections 3.1 and 3.2 line planning with travel quality and cost objective (LPQC).

The following observation from [Sch14] will be useful in the remainder of this paper:

**Observation 3.1.** Given a route set \( \mathcal{R} \) we can easily determine a corresponding line concept \( f(\mathcal{R}) \) by setting \( f_l(\mathcal{R}) := \max_{e \in \mathcal{E}} \left\lceil \frac{x_{e,l}(\mathcal{R})}{y} \right\rceil \).

The pair \( (\mathcal{R}, f(\mathcal{R})) \) is feasible by definition. In [Sch14] it was shown that it has minimal cost assuming the cost model \( \text{cost}(\mathcal{R}, f) := \sum_{l \in \mathcal{L}} \gamma_2 \cdot k_2^l \cdot f_l \). However, it is easy to see that given \( \mathcal{R} \) this line concept is cost-optimal for any reasonable cost model, in particular for the ones we will consider later.

### 3.3 Basics from game theory

In this section we describe some basic concepts from game theory which are used in the remainder of this paper. See, e.g., [NRTV07] for details.

Game theory studies the dynamics of situations where players try to minimize individual, conflicting objective functions. In a game \( (\mathcal{Q}, \mathcal{S}, g) \), each player \( q \in \mathcal{Q} \) has a set of strategies \( S_q \) among which he can choose. The individual objective function \( g_q(S_q, S^{\neq q}) \) of player \( q \) depends on his chosen strategy \( S_q \), but also on the strategies \( S^{\neq q} = (S_1, S_2, \ldots, S_{q-1}, S_{q+1}, \ldots, S_{|\mathcal{Q}|}) \) chosen by the other players.

A central concept of game theory is the concept of equilibria. A set of strategies \( (S_1, \ldots, S_{|\mathcal{Q}|}) \) is called (Nash) equilibrium if none of the players can improve his individual objective function by changing his strategy (given that all other players do not change their strategies). I.e., \( \hat{S} = (\hat{S}_1, \ldots, \hat{S}_|\mathcal{Q}|) \) is an equilibrium if for all \( q \in \mathcal{Q} \) it holds that
\[ g_q(\hat{S}_q, \hat{S}^{\neq q}) \leq g_q(S_q, \hat{S}^{\neq q}) \quad \forall S_q \in S_q. \]

Not all games have equilibria, and even if equilibria exist, they can be hard to find and they do not need to be unique.

A special class of games with good properties is the class of potential games. We call a function \( \Phi : \mathcal{S} = S_1 \times S_2 \times \ldots \times S_{|\mathcal{Q}|} \rightarrow \mathbb{R} \) potential function, if it satisfies the relation
\[ \Phi(S) - \Phi(S') = g_q(S_q, S^{\neq q}) - g_q(S'_q, S^{\neq q}) \] (4)
for all solutions \( S = (S_1, \ldots, S_{|\mathcal{Q}|}) \in \mathcal{S} \), all players \( q \in \mathcal{Q} \) and all solutions \( S' = (S_1, \ldots, S_{q-1}, S'_q, S_{q+1}, S_{|\mathcal{Q}|}) \) which can be obtained from \( S \) by exchanging the strategy of player \( q \). A game with potential function is called potential game. The existence of a potential function allows us to interpret the problem of finding an equilibrium to \( (\mathcal{Q}, \mathcal{S}, g) \) as an optimization problem. As we can easily verify in \( (4) \), an optimal solution to \( \Phi \) is an equilibrium for the considered game (although there may be equilibria which are not optimal for \( \Phi \)).

Furthermore, the relation \( (4) \) implies that every time a player changes his strategy to improve his personal objective (while the other players’ strategies remain unchanged), the solution becomes better with respect to \( \Phi \) and, in this sense, closer to an equilibrium. This motivates the approach of using best-response algorithms to find equilibria: in every step, one of the players changes his strategy to the best response with respect to the other players’ strategies, i.e., he picks a solution of the optimization problem \( \min_{S \in \mathcal{S}_q} g_q(S_q, S^{\neq q}) \) as a new strategy. If there is only a finite number of strategies, this procedure converges to an optimum of \( \Phi \), and hence to an equilibrium of the game in a finite number of steps.
A centralized way to evaluate a solution $S = (S_1, \ldots, S_{|Q|})$ is to sum up the individual objective functions to a centralized objective function $G(S) = \sum_{q \in Q} g_q(S_q, S^{-q})$. We call $S \in S$ system-optimal if it minimizes $G$.

There exist different concepts to measure the inefficiency of equilibria with respect to the centralized objective. The price of anarchy is defined as

$$\max_{S^* \text{ is an equilibrium}} \frac{G(S^*)}{\min_S G(S)}$$

Assuming that over time, selfish behavior will converge to equilibrium solutions, the price of anarchy gives a worst-case bound on the quality of such a convergence process.

The price of stability, $\min_{S^* \text{ is an equilibrium}} \frac{G(S^*)}{\min_S G(S)}$, in contrast, quantifies how far the best equilibrium (i.e.: the best solution that would be accepted by the players) is away from system optimality.

### 3.4 The line planning routing game

Our approach to finding good line planning solutions is to determine passengers’ routes $R$ in $G$ in such a way that $(R, f(R))$ is a good solution to the line planning problem.

To this end, we model the line planning problem as a routing game. The passengers $Q$ are the players. The strategies of a passenger $q$ are the routes $R_q$ from $(0, u_q)$ to $(v_q, 0)$. Each passenger has an individual objective function $g_q(R_q, R^{-q})$ which depends on his chosen route $R_q$ and the routes chosen by the other passengers $R^{-q}$. We call this game line planning routing game (LPRG). Equilibria $R^*$ of this game will be interpreted as solutions $(R^*, f(R^*))$ of the line planning problem.

The choice of the individual objective functions $g_q$ is of course crucial for the quality of the obtained solutions. Ideally, the individual objective functions should

- account for individual travel quality as well as costs in order to find a solution which is balanced between the two partly contradicting objectives of minimizing costs while maximizing quality
- model passengers’ behavior as realistically as possible.

We propose the following general model. The passengers objective is composed of the quality of the solution and a share of the overall costs as defined in (1):

$$g_q(R_q, R^{-q}) := \text{travel}_q(R_q, R^{-q}) + \text{transfer}_q(R_q) + \text{cost}_q(R_q, R^{-q})$$

To share the costs among the passengers, we equally divide the cost of all lines among all users

$$\text{cost}_q(R_q, R^{-q}) := \sum_{l \in R_q} \frac{\text{cost}(f(R))}{|\{q' \in Q : l \in R_{q'}\}|}$$

(this is referred to as line-based cost model in the following) or assign the line costs to the edges $e \in l$ as edge costs $\text{cost}_{(e,l)}$ and set

$$\text{cost}_q(R_q, R^{-q}) := \sum_{(e,l) \in R_q} \frac{\text{cost}_{(e,l)}(f(R))}{x_{(e,l)}(R_q, R^{-q})}$$

(called edge-based model in the following). If $\gamma_2 = 0$, $\text{cost}_{(e,l)}(f(R))$ is constant and we write $\text{cost}_{(e,l)} := \text{cost}_{(e,l)}(f(R))$. 

6
Note that in the definition of the quality functions in Section 3.1 and the individual objective functions in the section, we implicitly assumed that the passenger all have the same perception of quality of a travel route since we assume the weighting factors $\alpha_i$, $\beta_i$, and $\gamma_i = 1, 2$ to be the same for each passenger. It would be possible to replace these common weighting factors by a set of individual weighting factors for each passenger. However, for the sake of simplicity, in this paper we only consider the case of common weighting factors for all passengers.

3.5 Relation between LPQC and LPRG

In the line-based cost model it holds that

$$\sum_{q \in Q} \text{cost}_q(R_q, R^{-q}) = \text{cost}(f(R)).$$

This is also true in the edge-based cost model, as long as it is ensured that a line does not cross an edge completely empty.

Furthermore, by definition

$$\sum_{q \in Q} \text{travel}_q(R_q, R^{-q}) = \text{travel}(f(R), R) \quad \text{and} \quad \sum_{q \in Q} \text{transfer}_q(R_q) = \text{transfer}(R).$$

We obtain

$$\sum_{q \in Q} g_q(R_q, R^{-q}) = G(R, f(R)),$$

which implies that a system-optimal routing corresponds to an optimal solution of the line planning problem described in Sections 3.1 and 3.2.

Hence, on the one hand, if the price of anarchy in the LPRG is small, an equilibrium $R^*$ of the game provides us with a good approximation $(f(R^*), R^*)$ for LPQC.

Lemma 3.2. Denote by $I$ an instance of the LPQC. Assume that the price of anarchy for the corresponding instance $I_{RG}$ of LPRG is bounded by $\xi$. Then any equilibrium $R^*$ of $I_{RG}$ is a $\xi$-approximation $(f(R^*), R^*)$ for $I$.

So, on the one hand, finding an equilibrium to LPRG may be regarded as a new, decentralized, way of solving LPQC.

On the other hand, one may argue that in some cases, optimal solutions to LPQC are not desirable in practice. Indeed, it may happen that the routing $R$ in a solution $(f, R)$ to LPQC allots very long routes to some passengers for the "greater good" of a solution which is optimal with respect to objective (3).

As an example, consider the situation shown in Figure 1. There are two lines (depicted by gray arrows) from station $v_1$ to station $v_7$. One is a fast line which stops only at one intermediate station, the other one is a regional line which serves a geographically different route and visits many small stations in between. Assume that the transportation capacity of each line is $B = 100$. There are 100 passengers who want to travel from $v_1$ to $v_7$, 50 who want to travel from $v_2$ to $v_7$, and some smaller amounts of passengers traveling to and from the regional stations. Hence, both lines have to be established. Now, if the cost factors $\gamma_1$ and $\gamma_2$ in the objective function (3) are comparatively big, both lines will be established with frequency 1 in an optimal solution $(f', R')$ to LPQC. This means that 50 of the 100 passengers from $v_1$ to $v_2$ will be sent via the regional train route in an optimal solution.

However, if this solution was implemented in real life, at station $v_1$, when the passengers from $v_1$ and $v_7$ have to make a decision which train to board, the fast train is still empty. To implement the solution $(f', R')$ into practice, somebody would have to convince these 50 passengers to use a slower connection to reserve the seats in the fast train for the passengers from $v_2$ to $v_7$ boarding later. It is not hard to imagine, that the passengers from $v_1$ to $v_7$ would board the train anyway so that the starting in $v_2$ could not board or the train would be overcrowded.
This would not happen in the solution \((f(R^*), R^*)\) provided by an equilibrium \(R^*\) of the corresponding routing game LPRG. In this solution, all passengers from \(v_1\) to \(v_7\) would choose the fast train and the planner would be forced to provide enough frequency here to avoid overcrowding - unless taking the slow line along would be cheap enough to be a favorable option for the passengers. Hence, if we assume that cost_q(R_q) is an estimate of the real costs that a passenger would pay on a route \(R_q\), the solution \((f(R^*), R^*)\) defined by an equilibrium \(R^*\) of LPRG models passenger behavior in a better way, provides better estimates of actual solution quality and helps to avoid overcrowding and is hence, from this perspective, preferable to the solution \((f', R')\) found by LPQC.

4 Finding good line planning solutions using best-response algorithms

The approach of this paper is to find line planning solutions by finding equilibria of LPRG. We justified this approach in the preceding section. In this section, we concentrate on the process of finding such equilibria.

4.1 The best-response algorithm

A popular method for finding equilibria are best-response algorithms (compare Section 3.3). The scheme of a best-response algorithm for LPRG is given in Algorithm 1.

Algorithm 1 Best response algorithm

Require: PTN, line pool, set of passengers \(Q\), individual cost functions \(g_q\), maximal number of iterations in \(\mathbb{N} \cup \infty\)

Ensure: A routing \(\hat{R}\)

Start with an empty routing (or with an arbitrary non-empty routing).

while there are improvements for the passengers and maximal number of iterations is not reached do

    for Passenger \(q \in Q\) do

        Calculate optimal passenger route \(R_q\) based on the route choices of the other passenger \(R^{−q}\) (routing problem)

    end for

end while

In the remainder of this paper we are going to discuss the following questions:

1. How can we solve the routing problem?
2. Does the best-response algorithm converge?
3. How good is the solution \((f(\hat{R}), \hat{R})\) to LPQC defined by the solution \(\hat{R}\) from Algorithm 1?
4.2 The routing problem
In every step of the (LPRG) we have to solve the following routing problem \( (RP_q) \) for passenger \( q \):

**Definition 4.1.** Given the PTN \( G \), the line pool \( L \), origin \( u_q \), destination \( v_q \) and individual objective function \( g_q \) for passenger \( q \), and routes \( R_{q'} \) for all passengers \( q' \in Q \setminus \{q\} \), the routing problem \( (RP_q) \) consists of finding a route \( R_q \) from \( u_q \) to \( v_q \) such that \( g_q(R_q, R^{−q}) \) is minimized.

Unfortunately, in general the routing problem which has to be solved in each iteration of the best-response algorithm is NP-hard. We see in Section 4.2.1 that there are two components which make the problem hard:

1. line-based costs (see Theorem 4.2),
2. and frequency-based transfer times (see Theorem 4.3).

On the contrary, if costs are assumed to be edge-based with \( \gamma_2 = 0 \) and transfer times are neglected, the problem becomes much better tractable, as we are going to discuss in Section 4.2.2. Section 4.5 discusses how to incorporate frequency-based transfer times in a heuristic way.

4.2.1 NP-hardness of the routing problem
For determining the complexity of our problems we will use reductions from the set covering decision problem (SCP), where an instance is given by a set of elements \( M = \{m_1, \ldots, m_n\} \), a set of subsets \( C \) with \( C \subseteq M \) and an integer \( K \in \mathbb{N} \). The problem is to find a subset \( C' \subseteq C \) such that \( \bigcup_{C \in C'} C \supseteq M \) and \( |C'| \leq K \).

We first show that the assumption of line-based costs leads to an NP-hard routing problem.

**Theorem 4.2.** The routing problem \( (4.1) \) with line-based cost model is NP-hard, even if transfer times and transfer penalties are not taken into account, i.e. if \( \beta = 0 \) and \( \gamma_2 = 0 \).

**Proof.** We show that (SCP) given by \( (M, C, K) \) can be reduced to the routing decision problem with line-based costs. Given an instance \( (M, C, K) \) of (SCP) we construct an instance of the routing problem as follows.

We create a node \( v_0 \) and for each \( m_i \in M \), \( i = 1, \ldots, n \) a node \( v_i \) and an edge \( e_i = (v_{i−1}, v_i) \). For all \( C \in C \) we create a line \( l_C \in L \) containing all edges \( \{e_1 : m_i \in C\} \) and additional edges to ensure that the lines are connected paths in the PTN. We set edge lengths to \( c(e) := 0 \) for all edges related to \( m \in M \) and to \( c(e) := K + 1 \) for all additional edges. Furthermore we set the costs to \( \text{cost}_l = 1 \) for all lines. We consider a passenger \( q \) who wants to travel from \( v_0 \) to \( v_n \). The parameters of the objective function are \( \alpha_1 = \gamma_1 = 1 \) and \( \alpha_2 = \beta = \gamma_2 = 0 \). (Note that the same construction is possible if \( \gamma_1 = 0 \) and \( \gamma_2 = 0 \).) Now there is a solution to the routing problem with objective value lower or equal to \( K \) if and only if there is a solution to (SCP) with objective value lower or equal to \( K \).

Let \( C' \) be a solution to (SCP). Then the set of lines \( L' := \{l_C : C \in C'\} \) has costs \( \leq K \) and allows \( q \) to travel from origin to destination with \( 0 \) travel time. On the other hand, in every solution to the constructed instance of the routing problem with travel time \( \leq K \), \( q \) uses the edge sequence \( (e_1, \ldots, e_n) \), because otherwise his travel time would be \( > K \). Hence, \( C' = \{C \in C : q \text{ uses } l_C\} \) is a solution to (SCP). \( \square \)

The following example illustrates this construction. Consider the instance of (SCP) given by

\[
\begin{align*}
M &= \{1, 2, 3, 4\}, \\
C &= \{C_1 = \{1, 2\}, C_2 = \{1, 3\}, C_3 = \{3, 4\}\}, \\
K &= 2.
\end{align*}
\]

This leads to the PTN...
where $e_1, \ldots, e_4$ correspond to $\mathcal{M}$ and $e_5$ is a helping edge for $C_2$ with

$$c(e_i) = \begin{cases} 
0 & , i = 1, \ldots, 4 \\
3 & , \text{otherwise}
\end{cases}$$

and

$$\mathcal{L}_0 = \{l_1 = (e_1, e_2), l_2 = (e_1, e_5, e_3), l_3 = (e_3, e_4)\}.$$ 

Due to the result of Theorem 4.2, in the remainder of this paper we restrict to edge-based cost functions. However, even without considering costs, a similar hardness result can be proven for the routing problem with frequency-based transfer times:

**Theorem 4.3.** The routing problem (4.1) is NP-hard, even if transfer penalties and operational costs are not taken into account, i.e., $\beta = 0$ and $\gamma_1 = \gamma_2 = 0$.

**Proof.** Similarly to the proof of Theorem 4.2 we prove this theorem by reduction from (SCP). Let $(\mathcal{M}, C, K)$ denote an instance of (SCP) with $n = |\mathcal{M}|$. We create a linear PTN consisting of two parts. The first part is used to ensure that at most $K$ sets are chosen from $\mathcal{C}$. The second part is similar to the construction in the proof of Theorem 4.2 and is used to determine whether the chosen sets cover $\mathcal{M}$.

The first part of the PTN consists of vertices $v_i$ for $i = 1, \ldots, 2K + 1$ and edges $e_i = (v_i, v_{i+1})$, $i = 1, \ldots, 2K$ with $c(e_i) = 0$. Additionally there are edges $e_{2i-1} = (v_{2i-1}, v_{2i})$ with costs $K' + 1$, where $K' := \frac{2K+2n}{3}$. For every edge $e_{2i-1}$ with an odd index we introduce a line $l_{2i-1}$ which consists of this edge only.

The second part of the PTN consists of vertices $w_i$ for $i = 1, \ldots, 2n + 1$ and edges $a_i = (w_i, w_{i+1})$ for $i = 1, \ldots, 2n$ with $c(a_i) = 0$. Furthermore, we add edges $a_{ij}$ which connect all pairs of vertices $w_i$ and $w_j$ with $i < j$ and whose length is $c(a_{ij}) := K' + 1$. For each $i = 1, \ldots, n$ we introduce a line $l_{2i-1}$ which covers the edge $a_{2i-1}$. We connect both parts of the PTN by a transition edge $l = (v_{2K+1}, w_1)$.

For every $C \in \mathcal{C}$ we create a line $l_C \in \mathcal{L}$ containing all edges $\{e_{2i} : m_i \in C\}$ from the first part of the PTN, the transition edge $l$, and the edges $a_{2i}$ with $m_i \in C$ from the second part of the PTN. We add additional edges with lengths $K' + 1$ whenever needed to ensure that the lines are connected paths in the PTN.

In contrast to the proof of Theorem 4.2, in this proof we have $|C| + 1$ passengers. Each passenger $q_C$ with $C \in \mathcal{C}$ has origin $v_1$ and destination $v_{2K+1}$ and his route $R_{q_C}$ is identical to line $l_C$ from $v_1$ to $v_{2K+1}$. The passenger $q$ for which we have to solve the routing problem has origin $v_1$ and destination $w_{2n+1}$. We set the capacity in each train to $B := 1$. For the objective function we use the parameters $\alpha_1 = \alpha_2 = 1$ and $\beta = \gamma_1 = \gamma_2 = 0$. Note that line costs can be set to arbitrary values, since $\gamma_1 = \gamma_2 = 0$.

We now show that there is a solution to the considered instance of (SCP) if and only if there is a solution $R_q$ to the routing problem $(RP_q)$ with individual objective value $g_q(R_q, R^{-q}) \leq K'$.

First note that any such route in the first part of the PTN will use the lines $l_i$ on edges with an odd index and some lines $l_i$ on the ones with an even index. Furthermore, for all of these paths the contribution from the first part of the PTN to the individual objective function is $\frac{2K}{3}$, since the length of every used edge is 0 and on each such path there is a transfer at each station between a line $l_i$ that is used by one passenger and a line $l$ that is used by two passengers. In the second part of the PTN, only edges $a_i$ can be used in a route $R_q$ with $g_q(R_q, R^{-q}) \leq K'$. Now consider the transfer costs of such a path in the second part of the PTN. At each node in the second part of the PTN a transfer has to take place. Thereby, transfer time is $\frac{1}{2}$ if passenger $q$ did not use line $l_j$ in the first part of the PTN, $\frac{1}{3}$ if he used it. Since there are $2n$ such transfers, any path with
individual objective value $\leq K'$ uses on edge $a_2$, a line that was already used in the first part of the PTN.

Due to the construction of the lines $l_C$, this means that if there is a route $R_q$ with $g_q(R_q, \mathcal{R}^{-q}) \leq K'$, for each element $m_i \in \mathcal{M}$ at least one line $l_C$ with $C \supseteq m_i$ is used in the first part of the PTN.

Since not more than $K$ such lines can be used in $R_q$, there must be a solution to the considered instance of (SCP).

On the other hand, if there is a solution $C' = \{C_1, \ldots, C_k\}$ with $k \leq K$ to the considered instance of (SCP), using line $l_{C_i}$ on edge $e_{2i}$ for $i = 1, \ldots, k$ (and arbitrary lines on $e_{2i}$ for $i = k + 1, \ldots, K$) allows the passenger to choose a path with transfer time $\frac{\alpha}{3}$ in the second part of the PTN and thus yields an individual objective value $\leq K'$.

\[ \square \]

### 4.2.2 Cases with polynomially solvable routing problem

A convenient way to represent line planning problems is the change-and-go network (CGN) $\mathcal{G} = (\mathcal{V}, \mathcal{A})$, which was first introduced in [SS06a]. The set of nodes consists of station nodes $\mathcal{V}_{\text{stat}} := \{(v, \text{board}) : v \in \mathcal{V}\} \cup \{(v, \text{alight}) : v \in \mathcal{V}\}$ and travel nodes $\mathcal{V}_{\text{trav}} := \{(v, l) : l \in \mathcal{L}, v \in \mathcal{I}\}$. The set of arcs is $\mathcal{A} := \mathcal{A}_{\text{OD}} \cup \mathcal{A}_{\text{trans}} \cup \mathcal{A}_{\text{line}}$ with

- line arcs $\mathcal{A}_{\text{line}} := \{(v, l) : l \in \mathcal{L}, e \in \mathcal{I}\}$ for each edge $e$ covered by a line $l$,
- transfer arcs $\mathcal{A}_{\text{trans}} := \{((v_1, l_1), (v_2, l_2)) : v \in \mathcal{V}, l_1 \ni v, l_2 \ni v\}$,
- and arcs for boarding and alighting

$\mathcal{A}_{\text{OD}} := \{((v, \text{board}), (v, l)) : l \in \mathcal{L}, v \in \mathcal{I}\} \cup \{((v, l), (v, \text{alight})) : l \in \mathcal{L}, v \in \mathcal{I}\}$.

Now every route $R_q$ for a passenger $q$ can be uniquely represented in $\mathcal{G}$ as a path $P_q$ from $(u_q, \text{board})$ to $(v_q, \text{alight})$ in $\mathcal{G}$.

For $a \in \mathcal{A}$ we denote by $x_a(R)$ the number of passengers, using arc $a$ of the CGN, i.e., $x_a(R) := \{q \in \mathcal{Q} : P_q \ni a\}$ where $P_q$ is the path in the CGN corresponding to $R_q$. To abbreviate, we sometimes omit the routing and use the notation $x_a := x_a(R)$.

Let us now assume that (given $\mathcal{R}^{-q}$) the objective value of a route $R_q$ is the sum of edge weights over all edges contained in the corresponding path $P_q$, i.e., that there are arc weights $w_a^q(\mathcal{R}^{-q}) \geq 0 \forall a \in \mathcal{A}$ such that

$$ g_q(R_q, \mathcal{R}^{-q}) = \sum_{a \in P_q} w_a^q(\mathcal{R}^{-q}). \tag{6} $$

This is the case if costs are edge-based with $\gamma_2 = 0$ and $\alpha_2 = 0$. Indeed, it is easy to check that the weights

$$ w_a^q(\mathcal{R}^{-q}) := \begin{cases} \alpha_1 (e) + \frac{\text{cost}_{(a,l)}(\mathcal{R}^{-q})}{\beta} & \text{if } a = (e, l) \in \mathcal{A}_{\text{drive}} \\ \beta & \text{if } a \in \mathcal{A}_{\text{trans}} \end{cases} $$

fulfill (6). In Section 4.5, different approaches to define arc weights are studied.

If edge weights of the form (6) can be found, we obtain the following lemma:

**Lemma 4.4.** Consider an instance of the routing problem defined by a network $\mathcal{G}$, a set of passengers $\mathcal{Q}$, origin $u_q$ and destination $v_q$ of a passenger $q$ with individual objective function $g_q$ and set of routes $\mathcal{R}^{-q}$ for all $q' \in \mathcal{Q} \setminus \{q\}$. If there are arc weights $w_a^q(\mathcal{R}^{-q})$ as defined in (6), $(RP_q)$ can be solved in polynomial time.

**Proof.** In this case, any shortest path from $(u_q, \text{board})$ to $(v_q, \text{alight})$ with respect to the edge weights $w_a(\mathcal{R}^{-q})$ is an optimal solution to $I$. Hence, we can find a solution using, e.g., Dijkstra’s algorithm.

\[ \square \]

Hence, in this case, we can use the best-response algorithm 1 with, e.g., Dijkstra’s algorithm in the routing step to search for an equilibrium of the LPRG.
4.3 Existence of equilibria and convergence of the best-response algorithm

In this section we study under which assumptions equilibria to the LPRG exist and can be found by the best-response algorithm. We start with an example which shows that in the general case the existence of an equilibrium is not guaranteed.

4.3.1 Non-existence of equilibria

The following example shows that for some instances of LPRG there are no equilibria. This also implies that the best-response algorithm does not converge in this case.

We consider the PTN from Figure 2 with 12 nodes and 18 edges. Every edge is served by one directed line (which contains only the edge). The capacity of a train is $B = 1$. There are three main passengers $q_1$ from $u_1$ to $v_1$, $q_2$ from $u_2$ to $v_2$, and $q_3$ from $u_3$ to $v_3$ and six sets of auxiliary passengers: $Q^j_i$ for $i = 1, \ldots, 3$ and $j = 1, 2$ contains $M$ passengers from $v_j^i$ to $v_i$ (where $M$ is a sufficiently large number, e.g., $M > 12$). We denote by $Q'$ the union of the auxiliary passengers.

We consider the objective functions $g_q(R) := \tau_q(R_q, R - q)$ which, for the sake of simplicity, takes only the transfer time into account, i.e., $\alpha = \beta = \gamma_1 = \gamma_2 = 0$.

Note that the auxiliary passengers only have one strategy, let $R'$ denote the set of these strategies.

Each of the main passengers $q_i$ has two different strategies: to take the route $R^1_i$ starting with edge $(u_i, v_1^i)$ or to take the route $R^2_i$ starting with edge $(u_i, v_2^i)$.

We now show that there does not exist an equilibrium in the described situation. Assume that $R$ is an equilibrium of the described line planning routing game. Denote by $R^j_i$ the strategy chosen by $q_i$. Without loss of generality, assume that $j_1 = 1$. Then

$$g_2(R^1_1, R^1_2, R^j_3, R') = \left\{\begin{array}{ll}
\frac{1}{1+2} + \frac{1}{2+1} + \frac{1}{2+M+1} = \frac{7}{6} + \frac{1}{M+1} & \text{if } j_3 = 1 \\
\frac{1}{1+1} + \frac{1}{1+1} + \frac{1}{1+1} = \frac{2}{3} & \text{if } j_3 = 2
\end{array}\right.$$

and

$$g_2(R^1_1, R^2_2, R^j_3, R') = \left\{\begin{array}{ll}
\frac{1}{1+1} + \frac{1}{1+1} + \frac{1}{1+1} = \frac{2}{3} & \text{if } j_3 = 1 \\
\frac{1}{1+1} + \frac{1}{1+1} + \frac{1}{1+1} = \frac{2}{3} & \text{if } j_3 = 2
\end{array}\right.$$

Since $R$ is an equilibrium, we conclude that $j_2 = 1$, i.e., $R^j_2 = R^1_2$.

Now

$$g_3(R^1_1, R^1_2, R^1_3, R') = \frac{1}{1+2} + \frac{1}{2+1} + \frac{1}{1+M+1} = \frac{4}{6} + \frac{1}{M+2}.$$
and
\[ g_3(R_1^1, R_2^1, R_3^1, R') = \frac{1}{1+1} + \frac{1}{1+2} + \frac{1}{2+M+1} = \frac{1}{6} + \frac{1}{M+3}. \]

Since \( R \) is an equilibrium, we conclude that \( j_3 = 1 \), i.e., \( R_3^2 = R_3^1 \).

Now we have a look at the strategies for \( q_1 \):
\[ g_1(R_1^1, R_2^1, R_3^1, R') = \frac{1}{1+1} + \frac{1}{1+2} + \frac{1}{2+M+1} = \frac{1}{6} + \frac{1}{M+3} \]
and
\[ g_1(R_2^2, R_1^1, R_3^1, R') = \frac{1}{1+2} + \frac{1}{2+1} + \frac{1}{1+M+1} = \frac{1}{6} + \frac{1}{M+2}. \]

Thus, \( g_1(R_1^1, R_2^1, R_3^1, R') > g_1(R_2^2, R_1^1, R_3^1, R') \). This is a contradiction to \( R_1^1 \) being part of an equilibrium.

Due to the symmetry of the construction of the instance, the assumption that \( R_2^1 \) is part of an equilibrium leads to a contradiction in the same way.

### 4.3.2 Line planning routing games with potential functions

In contrast to the example from Section 4.3.1 as we see in Lemma 4.5, existence of equilibria and convergence can be guaranteed if for every \( a \in \mathcal{A} \) there is an arc weight function \( \bar{w}_a : \mathbb{N} \to \mathbb{R} \) such that
\[ g_q(R_q, R_{-q}) = \sum_{a \in P_q} \bar{w}_a(x_a) \tag{7} \]
for every route \( R_q \) from \( u_q \) to \( v_q \) and its corresponding path \( P_q \) in the CGN.

In case of edge-based costs with \( \gamma_2 = 0 \) and \( \alpha_2 = 0 \), such arc weight functions are given by
\[ \bar{w}_a(x) := \begin{cases} \alpha_1 e(c) + \frac{\text{cost}(c,l)}{x} & \text{if } a = (e,l) \in \mathcal{A}_{\text{drive}} \\ \beta & \text{if } a \in \mathcal{A}_{\text{transfer}} \end{cases} \tag{8} \]

**Lemma 4.5.** Let \( I := (G, L, Q, \{g_q : q \in Q\}) \) be an instance of the LPRG such that are weight functions as specified in (7) exist. Then

1. \( \Phi(R) := \sum_{a \in \mathcal{A}} \sum_{i=1}^{x_a(R)} \bar{w}_a(i) \) is a potential function for \( I \),
2. there exists an equilibrium to \( I \),
3. the best-response algorithm converges to an equilibrium in a finite number of steps,
4. each of the steps can be executed in polynomial time.

**Proof.** This proof follows standard arguments for convergence of atomic routing games, compare, e.g., [Rou07].

1. Let \( R \) and \( R' \) be two routings. We denote with \( P_q \) and \( P_q' \) the corresponding paths for passenger \( q \) in the CGN and with \( x_a := x_a(R) \) and \( x_a' := x_a(R') \) the corresponding flows on edge \( a \) of the CGN. We first observe that
\[
\Phi(R_q, R_{-q}) - \Phi(R_q', R_{-q}) = \sum_{a \in P_q \setminus P_q'} \bar{w}_a(x_a) - \sum_{a \in P_q \setminus P_q'} \bar{w}_a(x_a') = g_q(R_q, R_{-q}) - g_q(R_q', R_{-q}),
\]
hence \( \Phi \) indeed is a potential function by (4).
2. Hence, every optimum of $\Phi$ is an equilibrium of the game. Since the number of solutions is finite, there exists at least one optimum of $\Phi$/equilibrium of $I$.

3. Since in each step of the best-response algorithm there is a non-zero improvement in the individual objective function and thus also in the potential function, and the number of solutions is bounded, the best-response algorithm converges to an optimum of $\Phi$ which is an equilibrium.

4. We set $w_q (R^{-q}) := \bar{w}_a(x_a(R^{-q}) + 1)$. Then

$$g_q(R_q, R^{-q}) = \sum_{a \in P_q} \bar{w}_a(x_a(R))$$

$$= \sum_{a \in P_q} \bar{w}_a(x_a(R^{-q}) + 1)$$

$$= \sum_{a \in P_q} w_q (R^{-q}).$$

The proposition follows from Lemma 4.4.

Thus, if objective functions fulfill the conditions of Lemma 4.5 an equilibrium will be obtained after a finite number of steps, this holds in particular for all line planning routing games with $\alpha_2 = \gamma_2 = 0$ and edge-based costs.

### 4.4 Equilibria and system optima

In this section we investigate the quality of the equilibria of LPRG by comparing them to system-optimal solution of the line planning routing game. First, we illustrate that the equilibria which are found by the best-response algorithm are not necessarily optimal with respect to the line planning problem defined in Section 3.1, even if convergence is guaranteed by Lemma 4.5.

To this end we consider the PTN $G$ and line pool $L$ associated to the CGN shown in Figure 3 with $c((v_1, v_2)) = c((v_1, v_3)) = 99$, $c(e) = 0$ and $\forall e \in E \setminus \{(v_1, v_2), (v_1, v_3)\}$. We set the edge-based line costs of edge $(v_1, v_3)$ in line $l_2$ to cost $((v_1, l_2), (v_3, l_2)) := 100$, all other line costs are 0. We
consider two passengers: \( q_1 \) wants to travel from \( v_1 \) to \( v_2 \) and \( q_2 \) wants to travel from \( v_1 \) to \( v_4 \), with individual objective functions \( g_q(R) = c(R^q) + \sum_{(e,l) \in R_q} \text{cost}_{(e,l)}(f(R)) \), i.e., \( \alpha_1 = \gamma_1 = 1 \), \( \beta = \alpha_2 = \gamma_2 = 0 \). (The same construction is possible if we switch the role of \( \gamma_1 \) and \( \gamma_2 \) and assign high enough train capacity).

There are two equilibria:

1. \( R' \): \( q_1 \) uses line 1 and \( q_2 \) uses line 3. Both of them have costs of 99, the social costs are 198.

2. \( R^* \): \( q_1 \) uses line 2 and 4, \( q_2 \) uses line 2 and 5. Both of them have costs of 50, the social costs are 100.

Clearly, the second equilibrium is preferable to the first one, since for both passengers the individual objective function are almost double in the first one. However, e.g., when starting with an empty solution, the best-response algorithm will find the first equilibrium.

In the following example we see that the optimal solution to the line planning problem is not necessarily an equilibrium of the corresponding routing game at all: Consider the PTN \( G \) and line pool \( L \) associated to the CGN shown in Figure 4. This PTN is the same as the one used in the previous example, except that edge lengths \( c((v_1, v_2)) = 32 \) and \( c((v_1, v_3)) = 49 \) are changed. Like in the previous example we have \( c(e) = 0 \) \( \forall e \in E \setminus \{(v_1, v_2), (v_1, v_3)\} \) and the edge-based line cost of \( l_2 \) on edge \((v_1, v_3)\) is 100, all other line costs are 0. For the objective function we choose the parameters \( \alpha_1 = \gamma_1 = 1 \), \( \beta = \alpha_2 = \gamma_2 = 0 \) (again, the same construction is possible for \( \gamma_1 = 0 \) and \( \gamma_2 = 1 \) if capacity is high enough). Figure 4 shows the resulting CGN where all nonzero arc weight functions are shown. Let there be three passengers. \( q_1 \) wants to travel from \( v_1 \) to \( v_2 \), \( q_2 \) wants to travel from \( v_1 \) to \( v_4 \), \( q_3 \) wants to travel from \( v_1 \) to \( v_3 \).

There is only one equilibrium: \( R' \): \( q_1 \) uses line 1, \( q_2 \) uses line 3, \( q_3 \) uses line 2. \( G(R', f(R')) = 181 \). The solution which is optimal for the line planning problem is \( R^* \): \( q_1 \) uses line 2 and 4, \( q_2 \) uses line 2 and 5, \( q_3 \) uses line 2, with overall objective value \( G(R^*, f(R^*)) = 100 \). By extending the example given in Figure 4 in a straight-forward way, we see that for instances with an unbounded number of passengers, the price of stability is not bounded for the considered games: for \( n \) passengers we can construct an instance with price of stability \( H_n = \sum_{i=1}^{n} \frac{1}{i} \).

However, we can bound the price of anarchy by the number of passengers if the arc weight functions (7) fulfill an additional property.

**Lemma 4.6.** If there exist non-increasing arc weight functions \( \bar{w}_n \) with \( \bar{w}_n(1) \leq x \cdot \bar{w}_n(x) \) for all \( x \in \mathbb{N} \), the price of anarchy in the LPRG is at most the number of passengers.
Proof. Let \( X := \{X_1, \ldots, X_n\} \) be a socially optimal solution and let \( R := \{R_1, \ldots, R_n\} \) be an equilibrium. Assume that \( G(R, f(R)) > |Q|G(X, f(X)) \). Then there is at least one passenger \( q \) with \( g_q(R) > |Q|g_q(X) \). For this passenger \( q \) it follows that

\[
g_q(X_q, R^{-q}) = \sum_{a \in X_q} \tilde{w}_a(x'_a) \leq \sum_{a \in X_q} \frac{1}{\beta} \tilde{w}_a \leq \sum_{a \in X_q} \hat{x}_a \tilde{w}_a(x_a) \leq \sum_{a \in X_q} |Q|\tilde{w}_a(x_a) < g_q(R_q, R^{-q}),
\]

where \( \hat{x}_a := x_a(X_q, R^{-q}) \) denotes the number of passengers on arc \( a \) in \( (X_q, R^{-q}) \) and \( x'_a := x_a(R) \) the number of passengers on arc \( a \) in \( R \). This is a contradiction to the assumption that \( R \) is an equilibrium.

Corollary 4.7. If edge-based cost functions with \( \gamma_2 = 0 \) are considered and \( \alpha_2 = 0 \), the price of anarchy is bounded by the number of passengers.

Proof. The functions given in (8) are non-increasing. Furthermore we have for \( x > 0 \)

\[
x\tilde{w}_a(x) = \begin{cases} x\alpha_1 c(a) + \text{cost}(a) & \text{if } a \in \mathcal{A}_{\text{drive}} \\ x\beta & \text{if } a \in \mathcal{A}_{\text{transfer}} \end{cases} \geq \begin{cases} \alpha_1 c(a) + \text{cost}(a) & \text{if } a \in \mathcal{A}_{\text{drive}} \\ \beta & \text{if } a \in \mathcal{A}_{\text{transfer}} \end{cases} = \tilde{w}_a(1).
\]

Hence, if we use the best-response algorithm as a heuristic for the line planning problem, we obtain a solution with approximation ratio \( \frac{1}{|Q|} \) where \( Q \) is the number of passengers.

To see that there are indeed instances \( I \) with a price of anarchy that equals \( |Q| \) consider the example given in Figure 3. If we set the travel costs on \( (v_1, v_2) \) and \( (v_1, v_4) \) to 100, \( R' \) and \( R \) are still both equilibria and the price of anarchy is 2. We can easily extend this construction to an arbitrary number of passengers.

However, convergence to an equilibrium may be slow. To find a polynomial-time approximation algorithm for the initial line planning problem, the next lemma is more helpful.

Lemma 4.8. If there exist non-increasing arc weight functions \( \tilde{w}_a \) with \( \tilde{w}_a(1) \leq x \cdot \tilde{w}_a(x) \) for all \( x \in \mathbb{N} \), given an empty state of the game, calculating the best response once for every passenger leads to a routing \( R \) with

\[
\frac{G(R, f(R))}{G(X, f(X))} \leq |Q|
\]

where \( X \) is a system-optimal solution.

Proof. Let \( Q = \{1, \ldots, n\} \) be the set of passengers and \( S^q \) for \( q = 1, \ldots, n \) the route combination after choosing the best response \( R_q \) for passenger \( q \), i.e., \( S^q = (R_1, R_2, \ldots, R_q, \emptyset, \emptyset, \ldots, \emptyset) \). Furthermore let \( X \) be the socially optimal solution, where the passengers choose the route \( X_q, q = 1, \ldots, n \) with corresponding paths \( Y_q \) in the CGN.

Since arc weight functions are non-increasing, it holds that \( g_q(S_n) \leq g_q(S_q) \). Since \( R_q \) is a best response to \( (R_1, R_2, \ldots, R_{q-1}, \emptyset, \emptyset, \ldots, \emptyset) \) we have

\[
g_q(S_q) = \sum_{a \in P_q} \tilde{w}_a(x_a(S^q)) = \sum_{a \in P_q} \tilde{w}_a(x_a(S^{q-1}) + 1) \leq \sum_{a \in Y_q} \tilde{w}_a(x_a(S^{q-1}) + 1)
\]  \hspace{1cm} (9)
where $P_q$ denotes the path in the CGN corresponding to $R_q$. With this, it holds

$$G(S^n, f(S^n)) = \sum_{q \in \mathcal{Q}} g_q(S^n) \leq \sum_{q \in \mathcal{Q}} g_q(S^q)$$

$$\leq \sum_{q \in \mathcal{Q}} \sum_{a \in Y_q} \bar{w}_a(x_a(S^q) + 1)$$

due to (9)

$$\leq \sum_{q \in \mathcal{Q}} \sum_{a \in Y_q} \bar{w}_a(1)$$

since $\bar{w}_a$ non-increasing

$$\leq \sum_{q \in \mathcal{Q}} \sum_{a \in Y_q} x_a(\mathcal{X}) \cdot \bar{w}_a(\mathcal{X})$$

since $\bar{w}_a(1) \leq x \cdot \bar{w}_a(x)$

$$\leq |\mathcal{Q}| \sum_{q \in \mathcal{Q}} \sum_{a \in Y_q} \bar{w}_a(x_a(\mathcal{X}))$$

$$= |\mathcal{Q}| \sum_{q \in \mathcal{Q}} g_q(\mathcal{X}^n) = |\mathcal{Q}| \cdot G(\mathcal{X}^n, f(\mathcal{X}^n)).$$

\[\square\]

**Corollary 4.9.** If there exist non-increasing arc weight functions $\bar{w}_a$ with $\bar{w}_a(1) \leq x \cdot \bar{w}_a(x)$ for all $x \in \mathbb{N}$, a solution $(\mathcal{R}, f)$ to the line planning problem with $\frac{G(\mathcal{R}, f)}{G(\mathcal{R}^*, f)} \leq |\mathcal{Q}|$ can be found in polynomial time.

As described above, we can show that this bound is tight, i.e., there are instances where the best-response algorithm can get stuck in an equilibrium whose objective value is $|\mathcal{Q}|$-times the optimal solution value.

### 4.5 Heuristic approaches to the routing problem

In the preceding Sections 4.2-4.4 we have seen that in order to achieve polynomial running time of the best-response algorithm, to be able to prove convergence to an equilibrium, and give bounds on the quality of an equilibrium, strong restrictions on the parameters of the objective function have to be imposed.

In this section we investigate heuristic approaches to the routing problem for the general case, i.e., for individual objective functions

$$g_q(R_q, \mathcal{R}^{-q}) = \text{travel}_q(R_q, \mathcal{R}^{-q}) + \text{transfer}_q(R_q) + \text{cost}_q(R_q, \mathcal{R}^{-q})$$

with edge-based costs

$$\text{cost}_q(R_q, \mathcal{R}^{-q}) := \sum_{(e,l) \in E_q} \frac{\text{cost}_{(e,l)}(f(R_q, \mathcal{R}^{-q}))}{x_{(e,l)}(R_q, \mathcal{R}^{-q})}.$$

In this general case, the routing problem is NP-hard (Theorem 4.3) and the best-response algorithm does not necessarily converge (see Section 4.3.1). To overcome these difficulties, in this section we simplify the transfer time function $\tau_q$ and the edge-based cost function $\text{cost}_q$.

#### 4.5.1 Auxiliary frequencies

The first approach is to replace the frequencies $f(\mathcal{R})$ by auxiliary frequencies $\tilde{f}(\mathcal{R}^{-q})$ when determining a route for passenger $q$. This small trick allows us to define arc weights in accordance to Lemma 4.4 and hence, to solve the routing problem using Dijkstra’s algorithm in the CGN.

Let $\mathcal{Q}$ be a set of passengers and let $\mathcal{R} = \{R_q : q \in \mathcal{Q}\}$ be a set of strategies represented by paths in the CGN. We call an edge $(e,l) \in A$ critical for $\mathcal{R}$ if one additional passenger on the edge would increase the frequency, i.e., if $x_{(e,l)}(\mathcal{R}) \equiv 0 \mod B$. A line $l \in L$ is critical for $\mathcal{R}$ if it contains an
edge which is critical for \( R \). In order to find a route, represented by a path \( P_q \) in the CGN, given the routes for all other passengers \( R^{-q} \), we define the auxiliary frequencies

\[
\tilde{f}_l(R^{-q}) := \begin{cases} f_l(R^{-q}) + 1 & \text{if } l \text{ is critical for } R^{-q} \\ f_l(R^{-q}) & \text{otherwise.} \end{cases}
\]

We observe that for every line \( l \) and every passenger \( q \in Q \),

\[
\tilde{f}_l(R^{-q}) \geq f_l(R) \geq f_l(R^{-q}).
\]

For all non-critical lines we even have equality. Plugging in the auxiliary frequencies into \( \tau_q \) we obtain an auxiliary transfer time function

\[
\tilde{\tau}_q(R) := \sum_{i=1}^{n-1} \frac{1}{\tilde{f}_l_i(R^{-q}) + \tilde{f}_{l_{i+1}}(R^{-q})}
\]

(where \( l_1, \ldots, l_n \) are the lines used in \( R_q \)) which underestimates the transfer times \( \tau(R) \) in a routing \( R \). To find an overestimating heuristic measure for transfer times, we can consider

\[
\tilde{\tau}_q(R) := \tau_q^{ub}(R^{-q}) = \frac{1}{\tilde{f}_l_i(R^{-q}) + \tilde{f}_{l_{i+1}}(R^{-q})}.
\]

Using the same approach, we can define overestimating auxiliary edge-based cost functions as

\[
\tilde{\text{cost}}_q^{ub}(R) := \sum_{(e,l) \in R_q} \frac{\text{cost}(e,l)(\tilde{f}(R^{-q}))}{x(e,l)(R_q, R^{-q})} \geq \text{cost}_q(R)
\]

and underestimating auxiliary edge-based cost functions

\[
\tilde{\text{cost}}_q^{lb}(R) := \text{cost}_q(R^{-q}) = \sum_{(e,l) \in R_q} \frac{\text{cost}(e,l)(f(R^{-q}))}{x(e,l)(R_q, R^{-q})} \leq \text{cost}_q(R_q).
\]

We define over- and underestimated versions of the individual objective functions

\[
\tilde{g}_q^{ub}(R_q, R^{-q}) := \alpha_1 c(R) + \alpha_2 \tilde{\tau}_q^{ub}(R^{-q}) + \text{transfer}_q(R_q) + \tilde{\text{cost}}_q^{ub}(R_q, R^{-q}),
\]

\[
\tilde{g}_q^{lb}(R_q, R^{-q}) := \alpha_1 c(R) + \alpha_2 \tilde{\tau}_q^{lb}(R^{-q}) + \text{transfer}_q(R_q) + \tilde{\text{cost}}_q^{lb}(R_q, R^{-q})
\]

and obtain

\[
\tilde{g}_q^{lb}(R_q, R^{-q}) \leq g_q(R_q, R^{-q}) \leq \tilde{g}_q^{ub}(R_q, R^{-q}).
\]

Given a passenger \( q \) and a set of strategies \( R^{-q} \) for the remaining passengers, the auxiliary frequencies allow us to define weights for the arcs in the CGN which depend only on the strategy choices of the remaining passengers \( R^{-q} \). This observation is summarized in the following lemma.

**Lemma 4.10.** For arc weights

\[
\tilde{\omega}_a^{ub}(R^{-q}) := \begin{cases} \alpha_1 c(e) + \frac{\text{cost}(e)(\tilde{f}(R^{-q}))}{x(e,R^{-q}) + 1} & \forall a = (e,l) \in A_{\text{drive}} \\ \alpha_1 c(e) + \frac{\text{cost}(e)(\tilde{f}(R^{-q}))}{x(e,R^{-q}) + 1} + \beta & \forall a = ((v,l),(v',l')) \end{cases}
\]

or

\[
\tilde{\omega}_a^{lb}(R^{-q}) := \begin{cases} \alpha_1 c(e) + \frac{\text{cost}(e)(f(R^{-q}))}{x(e,R^{-q}) + 1} & \forall a = (e,l) \in A_{\text{drive}} \\ \alpha_1 c(e) + \frac{\text{cost}(e)(f(R^{-q}))}{x(e,R^{-q}) + 1} + \beta & \forall a = ((v,l),(v',l')) \end{cases}
\]

we have

\[
\tilde{g}_q^{ub}(R_q, R^{-q}) = \sum_{a \in P_q} \tilde{\omega}_a^{ub}(R^{-q}) \quad \text{and} \quad \tilde{g}_q^{lb}(R_q, R^{-q}) = \sum_{a \in P_q} \tilde{\omega}_a^{lb}(R^{-q})
\]

(where \( P_q \) denotes the path in the CGN corresponding to \( R_q \)) and the routing problem can be solved in polynomial time.
Here, the last statement follows from Lemma 4.4.

However, the use of the auxiliary objective functions $\bar{g}_q$ does not guarantee the existence of an equilibrium: In fact, in the counter example shown in Section 4.3.1 we have $f_j(\mathcal{R}) = f_j(\mathcal{R}^{-q})$ for all choices of $q$ and $R_q$. Hence, this example also proves the non-existence of an equilibrium for objective functions $\bar{g}_q$.

### 4.5.2 Auxiliary arc weights

Since the heuristic from section 4.5.1 does not always lead to an equilibrium, we consider a further simplification which guarantees the existence of an equilibrium and the convergence of the best-response-algorithm.

Consider a set of passenger routes $\mathcal{R} = \{R_q : q \in Q\}$ in the CGN and a transfer edge $a = ((v, l), (v, l'))$. Then the frequency of $l$ and $l'$ respectively is at least $\left\lceil \frac{x_a(\mathcal{R})}{\bar{w}_a(x)} \right\rceil$, since at least all passengers changing from $l$ to $l'$ have to use $l$ and $l'$ respectively. Additionally all frequencies are at most $\left\lceil \frac{|Q|}{B} \right\rceil$ since no more than all passengers can use any given line. This leads to the following approximate arc weight functions:

\[
\bar{w}_a^{lb}(x) := \begin{cases} 
\alpha_1 c(e) + \frac{\gamma_1 k_1 + \gamma_2 k_2}{x} + \beta & \text{if } a = (e, l) \in A_{\text{drive}} \\
\frac{\alpha}{2\|\Delta t\|} + \beta & \text{if } a \in A_{\text{transfer}}
\end{cases}
\]

and

\[
\bar{w}_a^{ub}(x) := \begin{cases} 
\alpha_1 c(e) + \frac{\gamma_1 k_1 + \gamma_2 k_2}{x} + \beta & \text{if } a = (e, l) \in A_{\text{drive}} \\
\frac{\alpha}{2\|\Delta t\|} + \beta & \text{if } a \in A_{\text{transfer}}
\end{cases}
\]

With $\bar{g}_q^{lb}(R_q, \mathcal{R}^{-q}) := \sum_{a \in P_q} \bar{w}_a^{lb}(x_a)$ and $\bar{g}_q^{ub}(P_q, \mathcal{R}^{-q}) := \sum_{a \in P_q} \bar{w}_a^{ub}(x_a)$, (where $P_q$ is the path in the CGN corresponding to $R_q$) we obtain:

**Lemma 4.11.** For every passenger $p \in Q$ with route $R_q$ and $\mathcal{R} = (R_q, \mathcal{R}^{-q})$ we have

$$\bar{g}_q^{lb}(R_q, \mathcal{R}^{-q}) = \sum_{a \in P_q} \bar{w}_a^{lb}(x_a) \leq g_q(R_q) \leq \sum_{a \in P_q} \bar{w}_a^{ub}(x_a) = \bar{g}_q^{ub}(R_q, \mathcal{R}^{-q}).$$

From Lemma 4.4, Lemma 4.5, Lemma 4.6 and Lemma 4.8 we conclude:

**Lemma 4.12.** For individual objective functions $\bar{g}_q^{lb}$ and $\bar{g}_q^{ub}$, the routing step of the best-response algorithm can be executed in polynomial time using arc weights $\bar{w}_a^{lb}(x_a)$ or $\bar{w}_a^{ub}(x)$, respectively, in the CGN.

With respect to these objective functions equilibria exist, the best-response algorithm converges towards an equilibrium, and the price of anarchy is at most $|Q|$. Furthermore, given an empty state of the so-defined game, calculating the best response once for every passenger leads to a state with costs of at most number of passengers times the costs of the social optimal solution.

### 5 Experiments

#### 5.1 Test instances

In order to find out how our best-response algorithm behaves and to gain first insights on its performance, we tested our best-response algorithm it on the three small test instances illustrated in Figures 5-7. The figure on the left shows the PTN with travel times and the lines, on the right.
Figure 5: Description of instance I1

Figure 6: Description of instance I2

line costs and OD-matrix are given. I1, shown in Figure 5 is our smallest example consisting of only three stations, two lines, and three passengers. I2 and I3 are bigger, the former consists of six stations, five lines, and 49 passengers, the latter contains eight stations, eight lines, and 2620 passengers.

Additionally we tested the algorithm on two more realistic instances GOE1 and GOE2 which are taken from the LinTim toolbox [Lin14]. These instances are both based on the PTN shown in Figure 8 which is based on the bus-network in Göttingen, Germany. It consists of 257 stations and 548 edges. A line pool consisting of 44 lines was generated for these experiments. The instances GOE1 and GOE2 differ in the number of passengers. For our experiments we modified the original OD-data from LinTim toolbox which consisted of 460146 passengers on 66048 OD-pairs. For GOE2, the number of passengers are scaled down by the factor $\frac{1}{30}$ and rounded, which leads to 6321 passengers on 6114 OD-pairs. For GOE1, the same calculation is done with the factor $\frac{1}{50}$, leading to OD-data of 1943 passengers and as many OD-pairs.

5.2 Running time

Table 1 gives a short overview of some computational aspects for using auxiliary arc-weight $\bar{w}_{lb}$ described in Section 4.5.2 with parameter set $(\alpha_1/\alpha_2, \beta, \gamma_1/\gamma_2) = (0.5/0.5, 0.1, 0.1/0.1)$ for I1, I2 and I3 and $(\alpha_1/\alpha_2, \beta, \gamma_1/\gamma_2) = (0.5/0.01, 0.01, 0.01/0.01)$ for GOE1 and GOE2. The number of iterations for different parameter choices and for the other heuristics described in Sections 4.5.1

<table>
<thead>
<tr>
<th>parameters:</th>
<th>I1</th>
<th>I2</th>
<th>I3</th>
<th>GOE1</th>
<th>GOE2</th>
</tr>
</thead>
<tbody>
<tr>
<td># iterations</td>
<td>2</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>computation time per iteration (in seconds)</td>
<td>0.01</td>
<td>0.045</td>
<td>2.26</td>
<td>45.78</td>
<td>143.35</td>
</tr>
<tr>
<td>% pass. without transfer</td>
<td>100</td>
<td>99.1</td>
<td>99.44</td>
<td>14.49</td>
<td>6.4</td>
</tr>
</tbody>
</table>

Table 1: Characteristica of found solutions for different instances for individual cost function (11).
Figure 7: Description of instance I3

<table>
<thead>
<tr>
<th>line</th>
<th>cost</th>
<th>stations</th>
</tr>
</thead>
<tbody>
<tr>
<td>l_1</td>
<td>4</td>
<td>(v_1, v_3, v_6, v_7)</td>
</tr>
<tr>
<td>l_2</td>
<td>4</td>
<td>(v_2, v_3, v_6, v_9)</td>
</tr>
<tr>
<td>l_3</td>
<td>1.8</td>
<td>(v_4, v_5)</td>
</tr>
<tr>
<td>l_4</td>
<td>2.8</td>
<td>(v_3, v_6, v_9)</td>
</tr>
<tr>
<td>l_5</td>
<td>3</td>
<td>(v_2, v_3, v_4)</td>
</tr>
<tr>
<td>l_6</td>
<td>4</td>
<td>(v_3, v_6, v_5, v_4)</td>
</tr>
<tr>
<td>l_7</td>
<td>3.8</td>
<td>(v_3, v_4, v_5, v_6)</td>
</tr>
<tr>
<td>l_8</td>
<td>5.8</td>
<td>(v_1, v_3, v_4, v_5, v_6, v_7)</td>
</tr>
</tbody>
</table>

Figure 8: Underlying public transportation network for instances GOE1 and GOE2
We observe that the time needed for one iteration increases with increasing network size and in particular with increasing number of passengers. This is to be expected, since each iteration consists of finding one shortest path per passenger and a subsequent adaption of arc lengths in the network. Note that in instances GOE1 and GOE2, the solution time per passenger is roughly the same in both instances. Furthermore, it is interesting to see that the number of iterations did not increase with the problem size. This indicates that with respect to running time, the approach could also be applicable for bigger instances, in particular in conjunction with speed-up techniques for shortest paths.

The fourth row in Table 1 shows the percentage of passengers who can travel without a transfer from origin to destination. This number reveals the different levels of closeness to reality of the five instances. While in the smaller instances all or nearly all passengers can travel without transfers, this number drops dramatically for instance GOE1 and GOE2. Comparing GOE1 and GOE2, it is interesting to see that less passengers have to transfer in GOE2. This can be explained by the fact that in case of higher demand, more lines can be established and hence there are more choices for the passengers.

### 5.3 Comparison of the different heuristic approaches to include frequencies

In this section we compare the different heuristic approaches from Section 4.5 on the instances I3 and GOE2, using the parameter sets $(\alpha_1/\alpha_2, \beta, \gamma_1/\gamma_2) = (0.5/0.5, 0.5, 0.1/0.1)$ and $(\alpha_1/\alpha_2, \beta, \gamma_1/\gamma_2) = (0.5/0.01, 0.1, 0.01/0.01)$ respectively. The results are shown in Table 2 and Table 3, where auxiliary freq ub and auxiliary freq lb use the individual objective functions $\tilde{g}_q^{ub}$ and $\tilde{g}_q^{lb}$ described in Section 4.5.1, respectively, and aux. arc-weight ub and aux. arc-weight lb denote the heuristics based on auxiliary arc weights $\bar{w}_a^{ub}$ and $\bar{w}_a^{lb}$ as described in Section 4.5.2, respectively.

We compare the average travel time per passenger, split up into driving time and transfer time, the total number of transfers, the percentage of passengers without a transfer, the number of lines, the costs of the line concept, split up into fixed costs and costs depending on the frequency and the objective value (according to the given parameter set) as sum of the objective values of all passengers.

For the transfer time and travel time, we show both the exact value (calculated using $g$) and the value estimated by the heuristic individual objective functions (in brackets).

We observe that, as proven in Section 4.5, transfer times are overestimated in the heuristics auxiliary freq ub and aux. arc-weight ub, while they are underestimated in auxiliary freq lb and aux. arc-weight lb. As was to be expected, the estimated transfer time in the heuristics using auxiliary frequencies is much closer to the exact transfer time than in the heuristics using auxiliary arcs weights.

Comparing heuristics based on auxiliary frequencies and heuristics based on auxiliary arc weights

### Table 2: Comparison of different heuristics for instance I3, using parameter set (0.5/0.5, 0.5, 0.1/0.1).

<table>
<thead>
<tr>
<th># iterations</th>
<th>auxiliary freq ub</th>
<th>auxiliary freq lb</th>
<th>aux. arc-weight ub</th>
<th>aux. arc-weight lb</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>4</td>
<td>5</td>
<td>4</td>
<td>3</td>
</tr>
<tr>
<td>travel time</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>drive</td>
<td>16.85</td>
<td>16.79</td>
<td>16.85</td>
<td>16.5</td>
</tr>
<tr>
<td>transfer real(est.)</td>
<td>1.87(1.87)</td>
<td>1.92(1.9)</td>
<td>0.54(2.06)</td>
<td>0.72(0.1)</td>
</tr>
<tr>
<td>total real(est.)</td>
<td>18.72(18.72)</td>
<td>18.72(18.69)</td>
<td>17.4(18.92)</td>
<td>17.22(16.6)</td>
</tr>
</tbody>
</table>

# transfers

| # transfers | 1210 | 1530 | 340 | 460 |
| % pass. without transfer | 54.58 | 47.33 | 87.02 | 82.44 |

| lines |                   |                   |                    |                    |
| # lines | 7 | 9 | 8 | 11 |
| $\sum$ fixed line costs | 1884 | 2352 | 2016 | 2664 |
| $\sum$ freq line costs | 4840 | 5880 | 3800 | 4352 |
| $\sum$ line costs | 6724 | 8152 | 5816 | 7016 |
| objective value | 25559.7 | 25732.82 | 23542.43 | 23437.75 |
Table 3: Comparison of different heuristics for the frequencies for instance GOE2, using parameter set (0.5/0.01, 0.1, 0.01/0.01).

<table>
<thead>
<tr>
<th># iterations</th>
<th>auxiliary freq ub</th>
<th>auxiliary freq lb</th>
<th>edge-based ub</th>
<th>edge-based lb</th>
</tr>
</thead>
<tbody>
<tr>
<td>travel time</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>drive</td>
<td>49.55</td>
<td>50</td>
<td>42.23</td>
<td>41.91</td>
</tr>
<tr>
<td>transfer real(est.)</td>
<td>9.9(9.93)</td>
<td>9.9(9.89)</td>
<td>9.36(27.48)</td>
<td>9.51(0.88)</td>
</tr>
<tr>
<td>total real(est.)</td>
<td>59.45(59.48)</td>
<td>59.89(59.89)</td>
<td>51.59(59.71)</td>
<td>51.42(42.79)</td>
</tr>
<tr>
<td># transfers</td>
<td>24927</td>
<td>25672</td>
<td>15838</td>
<td>15688</td>
</tr>
<tr>
<td>% pass. without transfer</td>
<td>2.77</td>
<td>2.09</td>
<td>6.81</td>
<td>6.3</td>
</tr>
<tr>
<td>lines</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td># lines</td>
<td>37</td>
<td>38</td>
<td>39</td>
<td>39</td>
</tr>
<tr>
<td>∑ fixed line costs</td>
<td>2398710</td>
<td>2402736</td>
<td>2542826</td>
<td>2542826</td>
</tr>
<tr>
<td>∑ freq line costs</td>
<td>3476608</td>
<td>3649197</td>
<td>2671705</td>
<td>2672386</td>
</tr>
<tr>
<td>∑ line costs</td>
<td>5875418</td>
<td>6051393</td>
<td>5214331</td>
<td>5215012</td>
</tr>
<tr>
<td>objective value</td>
<td>204291.96</td>
<td>206111.89</td>
<td>185927.84</td>
<td>185237.34</td>
</tr>
</tbody>
</table>

Table 4: Comparison of different parameter sets in the individual objective functions using auxiliary arc weights $w_{lb}$ on instance I3.

<table>
<thead>
<tr>
<th>parameter changes wrt. (0.5/0.5, 0.5, 0.1/0.1)</th>
<th>no changes</th>
<th>$\alpha_1 = 17.5$, $\alpha_2 = 15$, $\alpha_2 = 0$</th>
</tr>
</thead>
<tbody>
<tr>
<td># iterations</td>
<td>3</td>
<td>2</td>
</tr>
<tr>
<td>travel time</td>
<td></td>
<td></td>
</tr>
<tr>
<td>drive</td>
<td>16.5</td>
<td>16.5</td>
</tr>
<tr>
<td>transfer real(est.)</td>
<td>0.72(0.1)</td>
<td>0(0)</td>
</tr>
<tr>
<td>total real(est.)</td>
<td>17.22(16.6)</td>
<td>16.5(16.5)</td>
</tr>
<tr>
<td># transfers</td>
<td>460</td>
<td>0</td>
</tr>
<tr>
<td>% pass. without transfer</td>
<td>82.44</td>
<td>100</td>
</tr>
<tr>
<td>lines</td>
<td></td>
<td></td>
</tr>
<tr>
<td># lines</td>
<td>11</td>
<td>11</td>
</tr>
<tr>
<td>∑ fixed line costs</td>
<td>2664</td>
<td>2664</td>
</tr>
<tr>
<td>∑ freq line costs</td>
<td>4352</td>
<td>4352</td>
</tr>
<tr>
<td>∑ line costs</td>
<td>7016</td>
<td>7016</td>
</tr>
<tr>
<td>objective value for (0.5/0.5, 0.5, 0.1/0.1)</td>
<td>23437.75</td>
<td>22264.91</td>
</tr>
</tbody>
</table>

among each other, we observe that the heuristics which overestimate transfer time and line costs, find solutions with less transfers and less lines. However, travel time is slightly higher in these cases.

A comparison of the objective values of the results provided by the different heuristics surprisingly shows that the simple heuristics using auxiliary arc weights provide better results than the heuristics using auxiliary frequencies for both the small instance I3 and the more realistic instance GOE2 - although the latter provide much better estimates of travel times. In particular, the heuristics based on auxiliary arc weights seem to find paths with a smaller overall number of transfers and a bigger percentage of direct travelers. This effect is a bit more pronounced for the heuristic aux. arc-weight ub which overestimates transfer times, but also clearly visible for aux. arc-weight lb.

5.4 Influence of the parameters

In order to investigate how the choice of parameters in the objective function influences the outcome of the best-response algorithm, in Table 4 and 5 we vary the parameters in the heuristic individual objective functions used in aux. arc-weight lb on instance 3.

We observe that the equilibrium the algorithm found for the given parameter set is not the best one for the objective function for this parameter set. Surprisingly, increasing the weighting parameter for travel-time ($\alpha_1$), transfer-time ($\alpha_2$) or transfers ($\beta$) each lead to an equilibrium that is better socially and reduces the number of transfers to 0.

Increasing the influence of the costs - regardless of considering fixed costs, frequency-based costs or both - leads to an socially considerable worse equilibrium which is on top of that more expensive than any other considered here. This undesirable property might be a general disadvantage of the heuristic used here.
The different results are also shown in Figures 9(a)-9(d), where the thickness of the line arcs corresponds to the number of passengers using it or the frequencies of the lines respectively. The direct edges correspond to edges in the PTN, the bent edges to edges in the CGN.

The first image in Figure 9(a) is the solution for the original parameter set. For Figure 9(b), the transfer time is taken into account more which leads to a higher usage of the long line 8, leading to fewer transfers in this instance. A similar picture can be derived by increasing $\beta$, i.e. the transfer penalty. In Figure 9(c) less lines are used, due to the increased cost influence in the solutions. A corresponding effect is the higher load on the remaining lines. Additionally one edge is not used anymore in this solution. Note that the solution is exactly the same regardless of which cost parameter is increased. For the last solution 9(d), only travel and transfer times are taken into account. This results in a very level picture regarding the usage of the different lines, probably because taking new lines into the solution does not cost more and lines with similar frequencies are nice for the transfer times.

6 Conclusions and further research

We presented a new approach to solve line planning by solving a routing game where the passengers are the players who aim at minimizing a weighted sum of their travel time, transfer penalties, and a cost share. Under strong assumptions on the objective function (transfer time is not taken into account and line costs can be assigned to edges and are independent of frequencies) equilibria of this game can be found using the described best-response algorithm. In case that the objective function does not fulfill these properties, applicability and convergence of the best-response approach can be achieved by a slight modification of the individual objective functions.

A logical next step will be to evaluate whether the line planning routing game, besides being an interesting object of study in itself, does indeed lead to a good heuristic for line planning. First, more experiments of the type presented in Section 5 on instances of realistic size (in particular also with respect to passenger numbers) may lead to more insights on the performance of the different approaches presented in Section 4.5. A positive effect of increasing passenger numbers is that the approximate frequencies $f(R-q)$ and $\tilde{f}(R-q)$ become better estimates of actual frequencies $f(R)$. However, in the current version of the best-response strategy, in each iteration a shortest path for each passenger has to be found, hence running time increases with increasing number of passengers. For large passenger numbers it may thus make sense to use flow equilibration techniques in the inner loop instead of shortest path computations for each individual passenger.

Second, line planning solutions obtained with the routing game approach should be compared to state-of-the-art exact and heuristic solution methods for line planning with respect to objective value, running time, and practicability of the found solution (in the sense of Section 3.5).
Figure 9: Dependence of equilibrium solution from choice of parameters using the auxiliary arc weights $\bar{w}_{lb}$ on instance I3.
While the terms for travel time and transfers are quite intuitive, many different choices are possible for the cost-sharing among passengers. It remains an interesting question how to divide operational costs among passengers such that, on the one hand, the algorithmic approach is still viable, and on the other hand, cost shares are comparable to real-world travel costs. Furthermore, it would be interesting to investigate whether the routing game approach can also be applied to other planning problems which can be considered integrated network design and routing problem like, e.g., timetabling or delay management with integrated routing.

References


<table>
<thead>
<tr>
<th><strong>ERIM Report Series</strong></th>
<th><strong>Research in Management</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td>Date of publication</td>
<td>2014–12–03</td>
</tr>
<tr>
<td>Version</td>
<td>03–12–2014</td>
</tr>
<tr>
<td>Number of pages</td>
<td>30</td>
</tr>
<tr>
<td>Persistent URL for paper</td>
<td><a href="http://hdl.handle.net/1765/77431">http://hdl.handle.net/1765/77431</a></td>
</tr>
<tr>
<td>Email address corresponding author</td>
<td><a href="mailto:schmidt2@rsm.nl">schmidt2@rsm.nl</a></td>
</tr>
</tbody>
</table>
| Address | Erasmus Research Institute of Management (ERIM)  
RSM Erasmus University / Erasmus School of Economics  
Erasmus University Rotterdam  
PO Box 1738  
3000 DR Rotterdam, The Netherlands  
Phone: +31104081182  
Fax: +31104089640  
Email: info@erim.eur.nl  
Internet: [http://www.erim.eur.nl](http://www.erim.eur.nl) |
| Availability | The ERIM Report Series is distributed through the following platforms:  
RePub, the EUR institutional repository  
Social Science Research Network (SSRN)  
Research Papers in Economics (RePEc) |
| Classifications | The electronic versions of the papers in the ERIM Report Series contain bibliographic metadata from the following classification systems:  
Library of Congress Classification (LCC)  
Journal of Economic Literature (JEL)  
ACM Computing Classification System  
Inspec Classification Scheme (ICS) |