# Line Planning with Minimal Traveling Time

Anita Schöbel and Susanne Scholl

Institut für Numerische und Angewandte Mathematik,
Georg-August-Universität Göttingen
Lotzestrasse 16-18, 37083 Göttingen, Germany
{schoebel}{sscholl}@math.uni-goettingen.de

**Abstract.** An important strategic element in the planning process of public transportation is the development of a line concept, i.e. to find a set of paths for operating lines on them. So far, most of the models in the literature aim to minimize the costs or to maximize the number of direct travelers. In this paper we present a new approach minimizing the travel times over all customers including penalties for the transfers needed. This approach maximizes the comfort of the passengers and will make the resulting timetable more reliable. To tackle our problem we present integer programming models and suggest a solution approach using Dantzig-Wolfe decomposition for solving the LP-relaxation. Numerical results of real-world instances are presented.

**Keywords.** Line planning, real-world problem, integer programming, Dantzig-Wolfe decomposition

## 1 Motivation and related literature

In the strategic planning process of a public transportation company one important step is to find a suitable line concept, i.e. to define the routes of the bus or railway lines. Given a public transportation network PTN $= (S, E)$ with its set of stations $S$ and its set of direct connections $E$, a *line* is defined as a path in this network. The *line concept* is the set of all lines offered by the public transportation company, together with their *frequencies*, where the frequency $f_l$ of a line $l$ contains the number of vehicles serving line $l$ within the planning period considered. The *frequency of an edge* $e$, on the other hand, is the number of vehicles running along the edge.

The line planning problem has been well studied in the literature. For an early contribution we refer to Dienst, see [1]. The many models given after this time can be roughly classified into the following two types. In a *cost-oriented approach* the goal is to find a line concept serving all customers and minimizing the costs for the public transportation company. The basic cost model has been suggested in Claessens et al., see [2], where a binary (linear) programming formulation has been given. A solution approach by branch and cut has been developed in [3]. In [4] an alternative formulation with integer variables has been proposed. In [5]

ATMOS 2005
5th Workshop on Algorithmic Methods and Models for Optimization of Railways
http://drops.dagstuhl.de/opus/volltexte/2006/660

Bussieck et al. present a fast solution approach combining nonlinear techniques with integer programming.

In [6] and [7] the authors get rid of the assumption that the passengers are assigned a priori for example by modal split to different types of trains. This is done by assigning a certain *type* to every node in the PTN, representing for example the size of the station.Then the type of a line determines the stations they pass. For example a line of type 1 stops at every station it passes, a line of type 2 will not halt at a station of type 1 but at every station of type 2 or higher. Several models, correctness and equivalence proofs are presented.

Recently, a fast heuristic variable fixing procedure which combines nonlinear techniques with integer programming is proposed in [5].

[6] presents a model that reconsiders the stations at which the trains stop for a given line plan. This model is used to determine the halting stations in such a way that the total travel time of passengers is minimized. Lagrangian relaxation is used to find lower bounds on this problem. Preprocessing and tree search techniques augment the efficiency of the branch&bound framework.

A second class of models are the *customer-oriented approaches*. In the *direct travelers approach* by Bussieck et al. [8] (see also [4]) the goal is to maximize the number of direct travelers (i.e. customers that need not change the line to reach their destination). As constraint, the number of vehicles running along an edge is restricted for each edge in the PTN, i.e. upper and lower bounds on the allowed frequencies on each edge are taken into account. The model maximizes the amount of one group of customers but without considering the remaining ones which might have very many transfers during their trips. It also does not take into account the travel times for the customers: Sometimes it is preferable to have a transfer but reach the destination earlier instead of sitting in the same line for the whole trip but having a large detour. This is done in recent models by [9,10,11,12] in which the goal is to design lines in such a way that the traveling time of the customers is minimized. The special case of locating one single line so as to maximize the number of passengers is treated in [13]. None of these models includes the number of transfers of customers in the objective function, which will be the basic feature of the model presented in this paper.

Another approach is to take into account that the behavior of the customers depends on the design of the lines. A first cost-oriented model including such demand changes was treated with simulated annealing in two diploma theses in cooperation with *Deutsche Bahn*, see [14,15]. Finally, we want to mention the work by Quak [16] in which lines are not taken out of a given line pool as done by all other publications mentioned here, but constructed from the scratch.

In our work we develop a new model which allows to sum over all travel times over all customers including penalties for the transfers needed. The first ideas for this model have been presented in [17]. Here, we also show how different frequencies of the lines can be taken into account. The remainder of the paper is organized as follows. In Section 2 we introduce the new line planning model, discuss its complexity in Section 3 and then describe and discuss five integer programming models in Section 4. We present two ways to solve the LP-relaxation, one based on Dantzig-Wolfe decomposition (see Section 5). Finally, we present numerical results based on a real-world application of German Rail (DB).

## 2 Basic definitions

A *public transportation network* is a finite, undirected graph PTN $= (S, E)$ with a node set $S$ representing stops or stations, and an edge set $E$, where each edge $\{u, v\}$ indicates that there exists a direct ride from station $u$ to station $v$ (i.e., a ride that does not pass any other station in between). For each edge $\{u, v\}$ we assume that the driving time $t_{uv}$ is known.

We assume the PTN as given and fixed. We further assume that a *line pool* $\mathcal{L}$ is given, consisting of a set of paths in the PTN. Each line $l \in \mathcal{L}$ is specified by a sequence of stations, or, equivalently, by a sequence of edges. Let $E(l)$ be the set of edges belonging to line $l$. Given a station $u \in S$ we furthermore define $\mathcal{L}(u) = \{l \in \mathcal{L} : u \in l\}$ as the set of all lines passing through $u$.

Moreover, let $\mathcal{R} \subseteq S \times S$ denote the set of all origin-destination pairs $(s, t)$ where $w_{st}$ is the number of customers wishing to travel from station $s$ to station $t$.

The line planning problem then is to choose a subset of lines $L \in \mathcal{L}$, together with their frequencies, which

- allows each customer to travel from its origin to its destination,
- is not too costly, and
- minimizes the "inconvenience" for the customers.

In the literature, the main customer-oriented approach dealing with the inconvenience of the customers is the approach of [4] (see also [8]) in which the number of direct travelers is maximized. In our paper, however, we deal with the sum of *all* transfers over all customers. On a first glance, the problem to minimize the number of transfers seems to be similar to maximizing the number of direct travelers, but it can easily be demonstrated that both models are in fact different.

Note that considering the number of transfers only would lead to solutions with very long lines, serving all origin-destination pairs directly but having large detours for the customers. To avoid this we determine not only a line concept, but also a path for each origin-destination pair and count the number of transfers **and** the length of the paths in the objective function. This is specified next.

Given a set of lines $L \subseteq \mathcal{L}$, a customer can travel from its origin $s$ to its destination $t$, if there exists an $s$-$t$-path $P$ in the PTN only using edges in $\{E(l) : l \in L\}$. The "inconvenience" of such a path is then approximated by the weighted sum of the traveling time along the path and the number of transfers, i.e.

$$\text{inconvenience}(P) = k_1 \text{Time}_P + k_2 \text{Transfers}_P.$$

On the other hand, the cost of the line concept $L \subseteq \mathcal{L}$ is calculated by adding the costs $C_l$ for each line $l \in L$, assuming that such costs $C_l$ are known beforehand.

The *line planning problem* hence is to find a feasible set of lines $L \subseteq \mathcal{L}$ together with a path $P$ for each origin-destination pair, such that the costs of the line concept do not exceed a given budget $B$ and such that the sum of all inconveniences over all paths is minimized.

Since the capacity of a vehicle is not arbitrarily large, we have to extend the basic problem to include frequencies of the lines. This makes sure that there are enough vehicles along each edge to transport all passengers. If each origin-destination pair can be served, the line concept is called *feasible*. We remark that often the number of vehicles running along the same edge is also bounded from above, e.g., for safety reasons.

## 3 Complexity Results

In this section we first show that the line planning problem as defined above is NP-hard, even in a very simple case, corresponding to $k_1 = 0$ in the above definition.

**Theorem 1.** *The line planning problem is NP-complete, even if*

- *we only count the number of transfers in the objective function,*
- *the* PTN *is a linear graph.*
- *all costs $C_l$ are equal to one.*

*Proof.* In the decision version, the line planning problem in the above case can be written as follows:

Given a graph PTN $=(S, E)$ with weights $c_e$ for each $e \in E$, origin-destination pairs $\mathcal{R}$, and a budget $B$, does there exist a feasible set of $B$ lines with less than $K$ transfers?

We reduce the set covering problem to the line planning problem: Given a set covering problem in its integer programming formulation

$$\min\{1_n x : Ax \geq 1_m, x \in \{0,1\}^n\}$$

with an 0-1 $m \times n$ matrix $A$, and $1_k \in I\!R^k$ the vector with a 1 in each component, we construct a line planning problem as follows:

We define the PTN as a linear graph with $2m$ nodes $S = \{s_1, t_1, s_2, t_2 \ldots, s_m, t_m\}$ and edges $E = \{(s_1, t_1), (t_1, s_2), (s_2, t_2), (t_2, s_3), \ldots, (s_m, t_m)\}$. We define an origin-destination pair for each row of $A$,

$$\mathcal{R} = \{(s_i, t_i) : i = 1, \ldots, m\}.$$

For column $j$ of $A$ we construct a line $l_j$ passing through nodes $s_i$ and $t_i$ if $a_{ij} = 1$.

As an example, Figure 1 shows the line planning problem obtained from a set covering problem with

$$A = \begin{pmatrix} 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 0 & 0 & 1 & 1 \\ 0 & 1 & 0 & 1 \\ 0 & 1 & 1 & 0 \\ 1 & 0 & 1 & 0 \end{pmatrix}$$

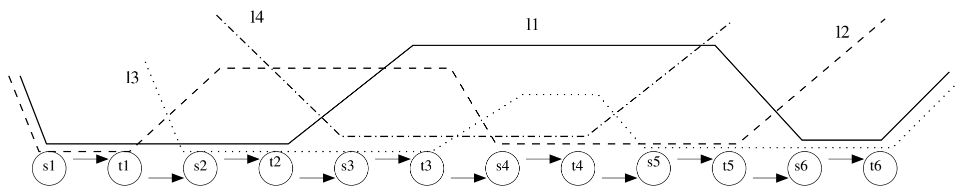


**Fig. 1.** Construction of the line planning problem in the proof of Theorem 1.

Setting $K = 0$ we hence have to show that a cover with less than $B$ elements exists if and only if the line planning problem has a solution in which all passengers can travel without changing lines. Due to our construction this is true and hence the theorem holds.

A question that might arise in this context, is what happens if the lines need not be chosen from a given line pool, but can be constructed as any path. Some of the basic cost models become very easy in this case, but unfortunately, the complexity status of the line planning problem treated in this paper does not change which can be shown by reduction to the Hamiltonian path problem (see [18]).

## 4 Models for the line planning problem

To model the line planning problem as integer program we use the PTN to construct a directed graph, the so-called *change&go network* $G_{CG} = (\mathcal{V}, \mathcal{E})$ as follows:

We extend the set $S$ of stations to a set $\mathcal{V}$ of nodes with nodes representing either station-line-pairs (change&go nodes: $\mathcal{V}_{CG}$) or the origins and destinations of the customers (origin-destination nodes: $\mathcal{V}_{OD}$), i.e. $\mathcal{V} := \mathcal{V}_{CG} \cup \mathcal{V}_{OD}$ with

- $\mathcal{V}_{CG} := \{(s, l) \in S \times \mathcal{L} : l \in \mathcal{L}(s)\}$ (set of all station-line-pairs)
- $\mathcal{V}_{OD} := \{(s, 0) : (s, t) \in \mathcal{R} \text{ or } (t, s) \in \mathcal{R}\}$ (origin-destination nodes)

The new set of edges $\mathcal{E}$ consists of directed edges between nodes of the same stations (representing that customers board or unboard a vehicle or change lines) and edges between nodes of the same line (representing the driving activities):

$$\mathcal{E} := \mathcal{E}_{change} \cup \mathcal{E}_{OD} \cup \mathcal{E}_{go}$$

with

- $\mathcal{E}_{change} := \{((s, l_1), (s, l_2)) \in \mathcal{V}_{CG} \times \mathcal{V}_{CG}\}$ (changing edges)
- $\mathcal{E}_l := \{((s, l), (s', l) \in \mathcal{V}_{CG} \times \mathcal{V}_{CG} : (s, s') \in E\}$ (driving edges of line $l \in \mathcal{L}$)
- $\mathcal{E}_{go} := \bigcup_{l \in \mathcal{L}} \mathcal{E}_l$ (driving edges)
- $\mathcal{E}_{OD} := \{((s, 0), (s, l)) \in \mathcal{V}_{OD} \times \mathcal{V}_{CG}$ and $((t, l), (t, 0)) \in \mathcal{V}_{CG} \times \mathcal{V}_{OD} : (s, t) \in \mathcal{R}\}$ (origin-destination edges)

We define weights on all edges $e \in \mathcal{E}$ of the change&go network representing the inconvenience customers have when using edge $e$. Given a set of lines $L \subseteq \mathcal{L}$ we then can determine the lines the customers are likely to use by calculating a shortest path in the change&go network for each single origin-destination pair. Therefore the choice of the edge costs $c_e$ is very important. We give two examples:

1. Customers only count transfers:

$$c_e = \begin{cases} 1 & : \quad e \in \mathcal{E}_{change} \\ 0 & : \quad \text{else} \end{cases}$$

   Note that in this case, it is possible to shrink the change&go network to a network with $|\mathcal{L}| + |S|$ nodes and $|\mathcal{E}_{change}| + |\mathcal{E}_{OD}|$ edges.
2. Real travel time:

$$c_e = \begin{cases} 0 & : \quad e \in \mathcal{E}_{OD} \\ \text{travel time in minutes} & : \quad e \in \mathcal{E}_{go} \\ \text{time needed for changing platform} & : \quad e \in \mathcal{E}_{change} \end{cases}$$

   More specific, to model the line planning problem as defined in Section 2, we set

$$c_e = \begin{cases} 0 & \text{if } e \in \mathcal{E}_{OD} \\ k_1 t_{uv} & \text{if } e = ((u, l), (v, l)) \in \mathcal{E}_{go} \\ k_2 & \text{if } e \in \mathcal{E}_{change} \end{cases}$$

Since we assume that customers behave selfish we need an implicit calculation of shortest paths (with respect to the weights $c_e$) within our model. This is obtained by solving the following network flow problem for each origin-destination pair $(s, t) \in \mathcal{R}$.

$$\theta x_{st} = b_{st},$$

where

- $\theta \in \mathbb{Z}^{|\mathcal{V}| \times |\mathcal{E}|}$ is the node-arc-incidence matrix of $G_{CG}$,

– $b_{st} \in \mathbb{Z}^{|\mathcal{V}|}$ is defined by

$$b^i_{st} = \begin{cases} 1 & : \quad i = (s,0) \\ -1 & : \quad i = (t,0) \\ 0 & : \quad \text{else} \end{cases}$$

– and $x^e_{st} \in \{0,1\}$ are the variables, where $x^e_{st} = 1$ if and only if edge $e$ is used on a shortest dipath from node $(s,0)$ to $(t,0)$ in $G_{CG}$.

To specify the lines in the line concept we introduce variables $y_l \in \{0,1\}$ for each line $l \in \mathcal{L}$, which are set to 1 if and only if line $l$ is chosen to be in the line concept. Our model, *Line Planning with Minimal Travel Times* (LPMT) can now be presented.

(LPMT1)

$$\min \sum_{(s,t)\in\mathcal{R}} \sum_{e\in\mathcal{E}} w_{st}\ c_e\ x^e_{st} \tag{1}$$

$$s.t. \quad \sum_{(s,t)\in\mathcal{R}} \sum_{e\in\mathcal{E}^l} x^e_{st} \le |\mathcal{R}||\mathcal{E}^l| y_l \qquad \forall\ l \in \mathcal{L} \tag{2}$$

$$\theta x_{st} = b_{st} \qquad \forall\ (s,t) \in \mathcal{R} \tag{3}$$

$$\sum_{l\in\mathcal{L}} C_l y_l \le B \tag{4}$$

$$x^e_{st}, y_l \in \{0,1\} \qquad \forall\ (s,t) \in \mathcal{R}, e \in \mathcal{E}, l \in \mathcal{L} \tag{5}$$

Constraint (2) makes sure that a line must be included in the line concept if the line is used by some origin-destination pair. Constraint (3) models the selfish behavior of the customers, i.e., that customers use shortest paths according to the weights $c_e$.

Having only constraints (2) and (3), the best line concept from a customer-oriented point of view would be to introduce all lines of the line pool. This is certainly no option for a public transportation company, since running a line is costly. Let $C_l$ be an estimation of the costs for running line $l$ and let $B$ be the budget the public transportation company is willing to spend. Then the budget constraint (4) takes the economic aspects into account.

The objective function we use is customer-oriented: We sum up the costs

$$\sum_{e\in\mathcal{E}} w_{st}\ c_e\ x^e_{st}$$

of a shortest path from $s$ to $t$ for each origin-destination pair $(s,t) \in \mathcal{R}$, i.e., we minimize the average costs of the customers.

We get three alternative formulations of this problem by substituting constraints (2) by one of the following constraints

$$\sum_{(s,t)\in\mathcal{R}} x^e_{st} \leq |\mathcal{R}| y_l \qquad \forall\, l \in \mathcal{L}, e \in \mathcal{E}^l \tag{6}$$

$$\sum_{e\in\mathcal{E}^l} x^e_{st} \leq |\mathcal{E}^l| y_l \qquad \forall\, l \in \mathcal{L}, (s,t) \in \mathcal{R} \tag{7}$$

$$x^e_{st} \leq y_l \qquad \forall\, (s,t) \in \mathcal{R}, e \in \mathcal{E}^l : l \in \mathcal{L} \tag{8}$$

We denote the formulation using constraints (6) (LPMT2), using (7) (LPMT3), and using (8) (LPMT4). As shown in [19], these formulations are equivalent, i.e. they are valid IP formulations for the same integer set $X$ of feasible solutions of the line planning problem. Nevertheless the bounds provided by the corresponding LP-relaxations differ. This will be analyzed next.

Let $X \subseteq \mathbb{Z}^n$ be a set of feasible solutions, and let two polyhedrons $P_A$ and $P_B$ be valid formulations for $X$, i.e., $X = P_A \cap \mathbb{Z}^n = P_B \cap \mathbb{Z}^n$. Then $P_A$. is said to be a *stronger* formulation than $P_B$ if $P_A \subset P_B$, see, e.g., [20]. In this case,

$$\min_{x\in X} cx \geq \min_{x\in P_A} cx \geq \min_{x\in P_B} cx,$$

i.e., the bound provided by the stronger formulation $P_A$ is better than the bound provided by $P_B$.

We can use this theory to analyze the strengthness of the four formulations presented for the line planning problem.

**Theorem 2.** *The convex hull of the integer set described by formulation (LPMT1) is denoted by $P_1$. The corresponding polyhedra described by formulation (LPMT2), (LPMT3), and (LPMT4) are denoted by $P_2$, $P_3$, and $P_4$, respectively. Then, the following holds:*

- *$P_4$ is stronger than $P_1$, $P_2$, and $P_3$.*
- *$P_3$ is stronger than $P_1$.*
- *$P_2$ is stronger than $P_1$.*
- *Comparing $P_3$ and $P2$, none of them is stronger than the other.*

The proof can be found in [19]. Note that in real world instances (LPMT3) comes out to be in most cases stronger than (LPMT2), see Section 5.1.

In (LPMT) we implicitly assume that all customers traveling from station $s$ to station $t$ choose the same path in the change&go network, i.e., the same set of lines. This can be done if edge capacities are neglected in (LPMT). In practice, this is usually not the case, since each vehicle only can transport a limited number of customers and usually there is only a limited number of vehicles possible along each line (e.g. due to safety rules). In the following, we therefore present an extension of (LPMT) taking into account the number of vehicles on each line in a given time period. Consequently, this formulation allows to split customers along different paths from $s$ to $t$ in the change&go network $G_{CG}$.

Let $N$ denote the capacity of a vehicle and let the new variables $f_l \in I\!N$ contain the frequency of line $l$, i.e., the number of vehicles running along line $l$ within a given time period. Furthermore we choose variables $x^e_{st} \in I\!N$ and change the vector $b_{st}$ to

$$b^i_{st} = \begin{cases} w_{st} & \text{if } i = (s,0) \\ -w_{st} & \text{if } i = (t,0) \\ 0 & \text{else} \end{cases}$$

Then the *Line Planning Model with minimal transfers and frequencies* (LPMTF) is the following:

(LPMTF)

$$\min \sum_{(s,t)\in\mathcal{R}} \sum_{e\in\mathcal{E}} c_e \; x^e_{st} \tag{9}$$

$$s.t. \quad \frac{1}{N} \sum_{(s,t)\in\mathcal{R}} x^e_{st} \le f_l \qquad \forall\ l \in \mathcal{L}, e \in \mathcal{E}_l \tag{10}$$

$$\theta x_{st} = b_{st} \qquad \forall\ (s,t) \in \mathcal{R} \tag{11}$$

$$\sum_{l\in\mathcal{L}} C_l f_l \le B \tag{12}$$

$$\sum_{l\in\mathcal{L}:k\in\mathcal{E}_l} f_l \le f^{max}_k \qquad \forall\ k \in E \tag{13}$$

$$x^e_{st}, f_l \in I\!N \qquad \forall\ (s,t) \in \mathcal{R}, e \in \mathcal{E}, l \in \mathcal{L} \tag{14}$$

Constraints (10) make sure that the frequency of a line is high enough to transport the passengers. If $f_l = 0$, the line $l$ is not chosen in the line concept. Constraints (11) are flow conservation constraints routing the passengers on the shortest possible paths. Note that the $x^e_{st}$ variables can take integer values, such that passengers may choose different paths for the same origin-destination pair. Constraint (12) is again the budget constraint but with costs for each vehicle of a line (which are multiplied by the frequency to get the costs of the line).The capacity constraint (13) may be included if upper bounds for the frequencies are present.

## 5 Solving the LP-relaxation

As we have shown in Section 3 the line planning problem is NP-hard, and, moreover in real-world instances, gets huge. But fortunately the formulations of (LPMT) and (LPMTF) have block diagonal structure with only few coupling constraints. Moreover, in both models, all blocks are totally unimodular since they represent network flow problems.

In Section 5.1 we identify cases in which the solution of the LP-relaxation can be found by solving shortest path problems. If this does not work we have to take advantage of the block diagonal structure by using a Dantzig-Wolfe decomposition, which is shown in Section 5.2.

### 5.1 Using the trivial solution

**Definition 1.** *A* trivial solution $(\bar{x}, \bar{y}^1)$, $(\bar{x}, \bar{y}^2)$, $(\bar{x}, \bar{y}^3)$, $(\bar{x}, \bar{y}^4)$ *of (LPMT1), (LPMT2), (LPMT3), (LPMT4), respectively, is defined as the solution* $\bar{x}^e_{st}$ *of the shortest path problems*

$$\theta x_{st} = b_{st} \quad \forall\ (s,t) \in \mathcal{R}$$

*on the change&go-network constructed of all lines of the line pool and*

$$\bar{y}^1_l := \frac{\sum_{(s,t)\in\mathcal{R}} \sum_{e\in\mathcal{E}^l} \bar{x}^e_{st}}{|\mathcal{E}^l||\mathcal{R}|} \qquad \forall\ l \in \mathcal{L} \quad \textit{(for (LPMT1))}$$

$$\bar{y}^2_l := \frac{\max_{e\in\mathcal{E}^l} \sum_{(s,t)\in\mathcal{R}} \bar{x}^e_{st}}{|\mathcal{R}|} \qquad \forall\ l \in \mathcal{L} \quad \textit{(for (LPMT2))}$$

$$\bar{y}^3_l := \frac{\max_{(s,t)\in\mathcal{R}} \sum_{e\in\mathcal{E}^l} \bar{x}^e_{st}}{|\mathcal{E}^l|} \qquad \forall\ l \in \mathcal{L} \quad \textit{(for (LPMT3))}$$

$$\bar{y}^4_l := \max_{(s,t)\in\mathcal{R}} \max_{e\in\mathcal{E}^l} \bar{x}^e_{st} \qquad \forall\ l \in \mathcal{L} \quad \textit{(for (LPMT4))}$$

*It is in general not unique and need not to be feasible in the sense that it fulfills the budget constraint.*

In real world instances it appears quite often that a trivial solution is an optimal solution of the LP-relaxation of (LPMT1). This is clear since the right hand sides $|\mathcal{R}||\mathcal{E}^l|$ of the coupling constraints (2) are chosen such that all passengers could use all edges of all lines. In real world only few edges of the network are used and so $K_l := \sum_{(s,t)\in\mathcal{R}} \sum_{e\in\mathcal{E}^l} x^e_{st}$ is much smaller than $|\mathcal{R}||\mathcal{E}^l|$, hence

$$\sum_{l\in\mathcal{L}} C_l \bar{y}^1_l = \sum_{l\in\mathcal{L}} C_l \frac{K_l}{|\mathcal{E}^l||\mathcal{R}|} \leq B$$

is often satisfied.

The following Lemma generalizes this for the other formulations. The proof can be found in [19].

**Lemma 1.** *Let* $i \in \{1,2,3,4\}$ *and let* $(\bar{x}, \bar{y}^i)$ *be a trivial solution of (LPMTi), as defined in Definition 1. If*

$$T_i := \sum_{l\in\mathcal{L}} C_l \bar{y}^i_l \leq B$$

*is satisfied, the trivial solution* $(\bar{x}, \bar{y}^i)$ *is an optimal solution of (LPMTi).*
*Note that for* $i = 4$ *the solution* $(\bar{x}, \bar{y}^4)$ *of the LP-relaxation of (LPMT4) is integer and thus if* $T_4 \leq B$ *holds, the trivial solution is an optimal solution to the original problem.*

In Table 1 we see the $T_i$-values for different line pool sizes, where the line costs are set to one. Note that in this case a value below one means that the trivial solution is always the optimal solution independently of the choice of the budget. Only if the given budget is smaller than the $T_i$ value, the trivial solution is not a feasible solution of the LP-relaxation of (LPMTi). Thus, table 1 demonstrates the difference of the strength of the formulations. The higher the $T_i$-value, the better the lower bound provided by the corresponding formulation.

We see that in real world instances the bound provided by (LPMT3) is much stronger than (LPMT2) even if we could not show this in general. This is due to the fact that there exists an instance in which (LPMT2) is stronger than (LPMT3) but in real world this hardly ever happens.

Regarding the $T_4$-values, we recall that in this formulation the $\bar{y}_l^4$ are integer valued and since all $C_l = 1$ this means that if we are allowed to choose more than $T_4$ lines out of the line pool, every passenger can travel on shortest path. If our budget is smaller, some passengers have a detour. In this case we have to use other methods to solve the problem like the Dantzig-Wolfe approach explained in the next section.

| No. | $\lvert\mathcal{L}\rvert$ | obj.val. | $T_1$ | $T_2$ | $T_3$ | $T_4$ |
|---|---|---|---|---|---|---|
| 1 | 10 | 2271.3 | 0.69 | 0.99 | 9.53 | 10 |
| 2 | 50 | 9459.9 | 0.20 | 0.35 | 25.31 | 48 |
| 3 | 100 | 24780.0 | 0.13 | 0.29 | 41.83 | 96 |
| 4 | 132 | 31654.2 | 0.11 | 0.26 | 53.12 | 129 |
| 5 | 200 | 15128.9 | 0.07 | 0.19 | 54.89 | 197 |
| 6 | 250 | 19096.0 | 0.05 | 0.16 | 61.07 | 235 |
| 7 | 275 | 20118.2 | 0.04 | 0.15 | 63.47 | 252 |
| 8 | 300 | 26598.3 | 0.06 | 0.19 | 72.35 | 282 |
| 9 | 330 | 26817.7 | 0.04 | 0.16 | 74.44 | 302 |
| 10 | 350 | 26450.0 | 0.07 | 0.23 | 90.04 | 331 |
| 11 | 375 | 27517.8 | 0.06 | 0.20 | 90.75 | 345 |
| 12 | 400 | 34781.3 | 0.06 | 0.20 | 100.05 | 370 |
| 13 | 423 | 35135.5 | 0.06 | 0.20 | 102.19 | 389 |

**Table 1.** Minimal budgets such that trivial solution is an optimal solution of the LP-relaxation of the different formulations of the (LPMT), see Lemma 1.

### 5.2 Using Dantzig-Wolfe decomposition

In this section we present an approach for solving the LP-relaxation of the (LPMT) formulations using Dantzig-Wolfe decomposition. The method can also be applied for solving (LPMTF) since the model structure is very similar. However, the numerical results deal with (LPMT). We will present two different decompositions. Since the blocks in both decompositions are totally unimodular, we know that the bound provided by the Master formulations is as good as the bound of the LP-relaxation (see [20]).

**One block for each origin-destination pair** (LPMT1(LP))

$$\min \sum_{(s,t)\in\mathcal{R}} \sum_{e\in\mathcal{E}} c^e_{st} x^e_{st} \tag{15}$$

$$\begin{array}{llll}
\multicolumn{4}{l}{\boxed{\begin{array}{l}\sum_{(s,t)\in\mathcal{R}} \sum_{e\in l} x^e_{st} \le |\mathcal{R}||\mathcal{E}^l| y_l \quad \forall\, l \in \mathcal{L} \\ \sum_{l\in\mathcal{L}} C_l y_l \le B\end{array}}} \quad \text{coupling constraints} \\
\boxed{X^{s_1,t_1}} & & & \\
& \boxed{X^{s_2,t_2}} & & \\
& & \ddots & \\
& & & \boxed{X^{s_r,t_r}}
\end{array} \left.\vphantom{\begin{array}{l} X \\ X \\ \ddots \\ X \end{array}}\right\} |\mathcal{R}| \text{ blocks}$$

where $X^{st} := \{x_{st} \in I\!R^{|\mathcal{E}|} : \theta x_{st} = b_{st}, 0 \le x^e_{st} \le 1,\ \forall\, e \in \mathcal{E}\}$
The coupling constraints can be written as

$$\begin{array}{r}-A_Y\ y + \sum_{(s,t)\in\mathcal{R}} A_X\ x_{st} \le 0 \\ Cy \le B\end{array}$$

where

- $A_X$ is an $|\mathcal{L}| \times |\mathcal{E}|$ matrix given by elements $a_{le} = 1$, if $e \in \mathcal{E}_l$, zero otherwise. It is equal for each origin-destination pair.
- $A_Y$ is an $|\mathcal{L}| \times |\mathcal{L}|$ diagonal matrix containing $|\mathcal{R}||\mathcal{E}^l|$ as its $l$th diagonal element.
- $C$ is the line cost vector $(C_1, \ldots, C_{|\mathcal{L}|})$.

So, we get the following coefficient matrix of (LPMT1):

$$\begin{pmatrix} -A_Y & A_X & \ldots & A_X \\ C & & & \\ & \theta & & \\ & & \ddots & \\ & & & \theta \end{pmatrix}$$

Defining the *weight-cost-parameters* $c^e_{st} := w_{st} c_e$, we get the following Master Problem corresponding to decomposition (15):

(Master 1)

$$
\begin{array}{ll}
z = \min \sum_{(s,t)\in\mathcal{R}} \sum_i (c_{st}\ x_{st}^{(i)})\alpha_{st}^i & \\
\quad s.t. \sum_{(s,t)\in\mathcal{R}} \sum_i (A_X\ x_{st}^{(i)})\alpha_{st}^i - A_Y y + Iv = 0 & \\
\quad \sum_{l\in\mathcal{L}} C_l y_l \leq B & \\
\quad \sum_i \alpha_{st}^i = 1 & \forall\ (s,t)\in\mathcal{R} \\
\quad y_l \geq 1 & \forall\ l\in\mathcal{L} \\
\quad v_l, \alpha_{st}^i, y_l \geq 0 &
\end{array}
$$

where the $|\mathcal{L}|$-vector $v$ are slack variables, and $x_{st}^{(i)}$ are the extreme points of $X^{st}$. This problem has $|\mathcal{L}|+1$ coupling constraints and $|\mathcal{R}|$ convexity constraints.

For each $(s,t)\in\mathcal{R}$ we obtain the following subproblem:

$$
\begin{array}{l}
z_{st} = \min\ (c_{st} - \pi A_X)x_{st} - \mu_{st} \\
\quad s.t.\ x_{st} \in X^{st}
\end{array}
$$

where $\{\pi_i\}_{i\in\mathcal{L}}$ are the dual variables of the coupling constraints, and $\{\mu_{st}\}_{(s,t)\in\mathcal{R}}$ are the dual variables of the convexity constraints.

The $X^{st}$ blocks correspond to shortest path problems which are known to be totally unimodular, hence the $x_{st}^{(i)}$-values are in $\{0,1\}^{|\mathcal{E}|}$. The formulations (LPMT2), (LPMT3), (LPMT4) as well as (LPMTF) can be reformulated analogously.

**One block for all origin-destination pairs** If we treat the $X^{st}$-blocks as one block we get the following reformulation:

(LPMT1(LP))

$$
\min \sum_{e\in\mathcal{E}} c^e x^e \tag{16}
$$

$$
\begin{array}{ll}
\boxed{\begin{array}{l} \sum_{e\in l} x^e \leq |\mathcal{R}||\mathcal{E}^l| y_l \quad \forall\ l\in\mathcal{L} \\ \sum_{l\in\mathcal{L}} C_l y_l \leq B \end{array}} & \text{coupling constraints} \\
\boxed{X} & \text{1 block}
\end{array}
$$

with $X := \{x \in I\!R^{|\mathcal{E}|} : x^e = \sum_{(s,t)\in\mathcal{R}} x_{st}^e\ \forall e\in\mathcal{E}, x_{st}\in X^{st}\}$ and $c^e := \sum_{(s,t)\in\mathcal{R}} c_{st}^e$.

The Master Program corresponding to decomposition (16) is

(Master 2)

$$
\begin{array}{l}
z = \min \sum_i (c\ x^{(i)})\alpha^i \\
\quad s.t. \sum_i (A_X\ x^{(i)})\alpha^i - A_Y\ y + Iv = 0 \\
\quad \sum_{l\in\mathcal{L}} C(l) y_l \leq B \\
\quad \sum_i \alpha^i = 1 \\
\quad v_l, \alpha^i, y_l \geq 0
\end{array}
$$

where the $|\mathcal{L}|$-vector $v$ are slack variables, and $x^{(i)}$ are the extreme points of $X$. This problem has $|\mathcal{L}|+1$ coupling constraints and one convexity constraints.

The subproblem of the $X$-block is

$$\begin{aligned} z = \min &\textstyle\sum_{(s,t)\in\mathcal{R}} (c_{st} - \pi A_X) x_{st} - \mu \\ \text{s.t. } & x_{st} \in X^{st} \end{aligned}$$

where $x^e := \sum_{(s,t)\in\mathcal{R}} x^e_{st}$ and $\{\pi_i\}_{i\in\mathcal{L}}$ are the dual variables of the coupling constraints, $\mu$ is the dual variable of the convexity constraint.

As in the previous formulation, the $x^{(i)}$-values are integer because they are the component wise sum over shortest path problem solution which are in $\{0, 1\}$. In this decomposition we loose the information of the exact paths of the customers which are needed in (LPMT3), (LPMT4) and (LPMTF) and thus this Master cannot be adapted to these formulations.

**Implementation** We implemented the Dantzig-Wolfe decomposition approach of (LPMT) using Xpress MP 2003 and Microsoft Visual C++ 6.0. The CPU times of this section are based on a 3.06 GHz Intel4 processor with 512 MB RAM. The subproblems where solved with Dijkstra´s shortest path algorithm.

In column 'CPU1' of table 2 we see the CPU times in seconds for solving the LP-relaxation of (LPMT1) using Dantzig-Wolfe approach with (Master2) for different line pool sizes of the network of German long distance trains. In column 'CPU2' we see the CPU times in seconds for solving the LP-relaxation of (LPMT3) using Dantzig-Wolfe approach with (Master1). We have mentioned that the lower bound provided by (LPMT3) is stronger than (LPMT1) and so the computation times increase in this case. We see, that using our approach it is possible to solve the LP-relaxation of (LPMT3) for medium sized networks within reasonable time. Note that the size of the problem not only depends on the size of the line pool but on the number of origin-destination pairs and the size of the PTN which may be much smaller e.g. in urban underground networks. Solving the LP-relaxation of the weaker (LPMT1) formulation is possible even for big real world instances like the long distance network of German railway within two and a half hours.

As we have seen, the main problem of our approach is the size of the change&go-network depending mainly on the size of the line pool. A wise choice of a possibly small line pool is therefore advisable. On the other hand it makes sense to analyze the underlying PTN. For example if two lines go parallel for a long time, it is sufficient to add changing edges only at the first and the last station. Also arcs between stations without changing possibility can be shrunken to decrease the size of the network.

| No. | $\lvert\mathcal{L}\rvert$ | $\lvert\mathcal{R}\rvert$ | CPU1 | CPU2 |
|---|---|---|---|---|
| 0 | 3 | 2 | 0.05 | 0.1 |
| 1 | 10 | 2602 | 1 | 228 |
| 2 | 50 | 4766 | 3 | 606 |
| 3 | 100 | 11219 | 16 | 8706 |
| 4 | 132 | 18238 | 48 | M |
| 5 | 200 | 10126 | 78 | M |
| 6 | 250 | 13246 | 329 | M |
| 7 | 275 | 14071 | 691 | M |
| 8 | 300 | 17507 | 1171 | M |
| 9 | 330 | 18433 | 1911 | M |
| 10 | 350 | 17095 | 1814 | M |
| 11 | 375 | 18350 | 2727 | M |
| 12 | 400 | 22191 | 4789 | M |
| 13 | 423 | 22756 | 8715 | M |

**Table 2.** CPU times of the LP-relaxation of (LPMT1) and (LPMT3) using Dantzig-Wolfe approach with (Master2) and (Master1), respectively, for different line pool sizes. M denotes "'out of memory"'.

## 6 Conclusions

We developed integer programming models for the line planning problem with the goal to minimize the travel times over all customers including penalties for the transfers needed and proposed an extension that includes frequencies. We showed that the problem is NP-hard. Since the problem gets huge, a straightforward solution of the LP relaxation is not possible. We showed that in many real world cases the trivial solution is optimal or, if it is infeasible, it can be found by a solution approach based on Dantzig-Wolfe decomposition. Computational results for various real world instances and different decompositions were presented.

We are currently working on a branch&price algorithm and heuristics to get an integer solution.

## References

1. Dienst, H.: Linienplanung im spurgeführten Personenverkehr mit Hilfe eines heuristischen Verfahrens. PhD thesis, Technische Universität Braunschweig (1978)
2. Zwaneveld, M., Claessens, M., van Dijk, N.: A new method to determine the cost optimal allocation of passenger lines. In: Defence or Attack: Proceedings 2nd TRAIL Phd Congress 1996, part 2, TRAIL Research School (1996)
3. Goossens, J., van Hoesel, C., Kroon, L.: A branch and cut approach for solving line planning problems. Transportation Science **38** (2004) 379–393
4. Bussieck, M.: Optimal lines in public transport. PhD thesis, Technische Universität Braunschweig (1998)

5. Bussieck, M., Lindner, T., Lübbecke, M.: A fast algorithm for near cost optimal line plans. Mathematical Methods of Operational Research **59** (2004) 205–220
6. Goossens, J.: Models and algorithms for railway line planning problems. PhD thesis, University of Maastricht (2004)
7. Goossens, J., van Hoesel, C., Kroon, L.: On solving multi-type line planning problems. Technical Report RM/02/009, University of Maastricht (2002) METEOR Research Memorandum.
8. Bussieck, M., Kreuzer, P., Zimmermann, U.: Optimal lines for railway systems. European Journal of Operational Research **96** (1996) 54–63
9. Borndörfer, R., Grötschel, M., Pfetsch, M.E.: Models for line planning in public transport. Technical Report 04-10, ZIP Berlin (2004)
10. Borndörfer, R., Grötschel, M., Pfetsch, M.E.: A path-based model for line planning in public transport. Technical Report 05-18, ZIP Berlin (2004)
11. Schneider, R.S.: A new customer-oriented cost model for the line planning problem. Master's thesis, Technische Universität Kaiserslautern (2005)
12. Laporte, G., Marin, A., Mesa, J., Ortega, F.: An integrated methodology for rapid transit network design. In: Algorithmic Methods for Railway Optimization. Lecture Notes on Computer Science. Springer (2004) presented at ATMOS 2004, to appear.
13. Laporte, G., Mesa, J., Ortega, F.: Maximizing trip coverage in the location of a single rapid transit alignment. Annals of Operations Research **136** (2005) 49–63
14. Klingele, H.: Verfahren zur Optimierung eines Linienkonzeptes der Deutschen Bahn AG. Master's thesis, Universität Karlsruhe (2000)
15. Schmidt, M.: Modelle zur Linienoptimierung im Zugverkehr unter Berücksichtigung der Nachfrage. Master's thesis, Universität Kaiserslautern (2001)
16. Quak, C.: Bus line planning. Master's thesis, TU Delft (2003)
17. Schöbel, A., Scholl, S.: Planung von Linien mit minimalen Umsteigevorgängen. In Mattfeld, D., ed.: Proceedings of the GOR-workshop on "Optimierung im öffentlichen Nahverkehr". (2003) 69–89 http://server3.winforms.phil.tu-bs.de/gor/tagung34/.
18. Schöbel, A., Scholl, S.: Line planning with minimal transfers. NAM Bericht 72, Georg-August Universität Göttingen (2004)
19. Scholl, S.: Customer-oriented line planning. PhD thesis, Technische Universität Kaiserslautern (2005)
20. Wolsey, L.: Integer Programming. John Wiley & Sons (1998)