Report in Wirtschaftsmathematik
Nr. 86/2003

Locating stops along bus or railway lines — a bicriterial problem

A. Schöbel

University of Kaiserslautern

January 7, 2003

Abstract

In this paper we consider the location of stops along the edges of an already existing public transportation network, as introduced in [SHLW02]. This can be the introduction of bus stops along some given bus routes, or of railway stations along the tracks in a railway network. The goal is to achieve a maximal covering of given demand points with a minimal number of stops. This bicriterial problem is in general NP-hard. We present a finite dominating set yielding an IP-formulation as a bicriterial set covering problem. We use this formulation to observe that along one single straight line the bicriterial stop location problem can be solved in polynomial time and present an efficient solution approach for this case. It can be used as the basis of an algorithm tackling real-world instances.

1 Introduction

In the design of a public transportation network, the number and the location of the stops (or stations) has to be planned carefully, see, e.g., [DAL82]. Unfortunately, it is not clear in advance, how many stops are reasonable,

and where they should be built. Even from a customer-oriented point of view, the following two conflicting effects of stops apply.

- On the one hand, many stops are advantageous, since they increase the accessibility for the customers. In bus transportation, it is often assumed that a customer will only use a bus, if the next bus stop is within a distance of at most 400 m. In rail transportation, this *covering radius* is larger, and is usually assumed to be 2 km.

- On the other hand, each additional stop increases the transportation time (e.g., by 2 minutes in rail transportation) for all trains or busses stopping there.

Moreover, this additional running time of the trains (or busses) is costly for the transportation company, and also fixed costs arise for establishing a new stop. Consequently, it makes sense to establish as few stops as possible in such a way, that all customers are covered. For a given finite set of possible new locations, this has been done in the *discrete* stop location problem which turns out to be an unweighted set covering problem (as tackled in [TSRB71]). In the context of stop location this set covering problem has been solved by [Mur01a] using the Lagrangian-based set covering heuristic of [CFT99], and applied in bus transportation in Brisbane, Australia, see [MDSF98, Mur01a, Mur01b]. Recently, another discrete stop location model has been developed by Laporte et al. [LMO02]. They investigate which candidate stops along one given line in Seville should be opened, taking into account constraints on the interstation space. The problem is solved by a longest path algorithm in an acyclic graph.

On the other hand, in the *continuous* stop location problem, the whole track system (or the routes of the busses) is allowed for locating stations. This problem is discussed in [SHLW02] under the objective of minimizing the number or costs of the new stations while covering all demand points. A similar covering model has been considered in [KPS$^+$02]. As part of a project with the largest German rail company (DB), another objective function has been developed in the continuous stop location problem, namely the minimization of the *door-to-door traveling time* over all customers, which is given by the access time of the customers to their closest station, their travel time within the vehicle and their time to go from the final station to their destination, see [HLS$^+$01]. An overview about continuous stop location is provided in [Sch02].
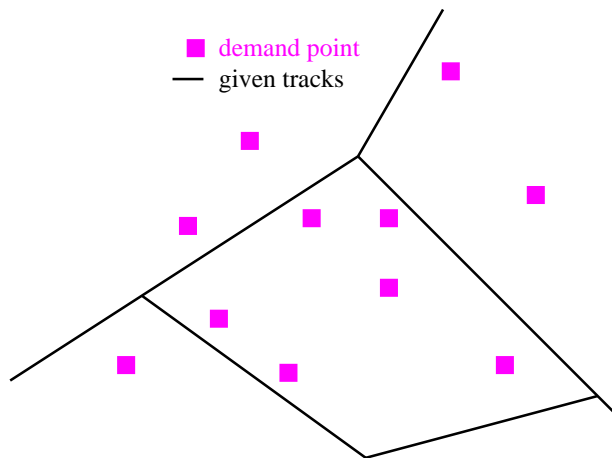
Figure 1: The set of tracks $\mathcal{T}$ and a set of demand points $\mathcal{D}$ in the plane.

In this paper we extend the continuous stop location problem as defined in [SHLW02] to a bicriterial problem. We need the following notation.

Let $\mathcal{D} \subseteq \mathbb{R}^2$ be a given finite set of demand points, and PTN $= (V, E)$ be the current public transportation network, given as a set of already existing stations or breakpoints V and their direct connections E. Then the set $\mathcal{T}$ of all points of the linear embedding of the graph PTN represents the given track system (for railways) or the bus routes (for bus transportation). Given a distance measure γ_d (which may depend on the demand point d), a demand point d is *covered* by a stop $s \in \mathcal{T}$, if $\gamma_d(d, s) \leq r$. In the following we assume that γ_d is a norm-distance for each demand point d. To allow different distance functions for each demand point is due to the possibly different environments close to the demand points and allows to model the distance functions more accurately. (Note that it is also possible to allow γ_d to be a distance derived from a gauge function. A gauge is defined similar to a norm, but without requiring symmetry, i.e., $\gamma_d(x, y) = \gamma_d(y, x)$ needs not be satisfied, see, e.g., [Min67].)

For a set $S \subseteq \mathcal{T}$ we define

$$\text{cover}(S) = \{d \in D : \text{ there exists } s \in S \text{ such that } \gamma_d(d, s) \leq r\}$$

as the *cover* of a set of stops S.

The goal of the (unweighted) *continuous stop location problem (CSL)* as defined in [SHLW02] is to find a set of (new) stops $S \subseteq \mathcal{T}$ with minimal cardinality, covering all demand points. This problem has been shown to be NP-complete.

Theorem 1 ([SHLW02]) *(CSL) is NP-complete.*

However, in a practical setting, one might not want to cover all demand points $\mathcal{D}$ but only a given percentage of the population. Hence let us assume that for each demand point, we have given a weight w_d representing the number of customers who would like to use public transportation, if the next station was closer than r. Then the function

$$f_{\text{cover}}(S) = \sum_{d \in \text{cover}(S)} w_d$$

gives the number of (potential) customers which live closer than r to some stop in S.

Certainly, it is preferable to cover as many customers as possible, i.e, to maximize $f_{\text{cover}}(S)$. On the other hand, establishing many new stops is costly and increases the travel time for the customers in the trains (or busses), because each stop needs an additional time of, e.g., two minutes. Since this causes dissatisfaction for the customers we use

$$f_{\text{cost}} = |S|$$

as a second objective function. The *bicriterial stop location problem (BSL)* can now be stated.

(BSL)

Given $G = (V, E)$ with its set of points of its planar embedding $\mathcal{T} = \bigcup_{e \in E} e \subseteq \mathbb{R}^2$, as well as a finite set of points $\mathcal{D} \subseteq \mathbb{R}^2$ with weights w_d and norms (or gauges) γ_d for all $d \in \mathcal{D}$, find a set $S \subseteq \mathcal{T}$ such that both

$$\begin{aligned} f_{\text{cost}} &= |S| \quad \text{and} \\ -f_{\text{cover}} &= - \sum_{d \in \text{cover}(S)} w_d \end{aligned}$$

are minimized.

2 Constraint problems and lexicographic minimality

What we mean by "minimizing both" objective functions is to find Pareto solutions of the problem with respect to f_{cost} and f_{cover}. Recall (e.g., from textbooks as [Ste89, Ehr00]) that if $S_1, S_2 \subseteq \mathcal{T}$ denote two feasible sets of stops, S_1 dominates S_2 if

$$f_{\text{cost}}(S_1) \leq f_{\text{cost}}(S_2) \text{ and}$$
$$f_{\text{cover}}(S_1) \geq f_{\text{cover}}(S_2),$$

where at least one of the inequalities is strict. Then a *Pareto solution* S^* is a feasible set of stops which is not dominated by any other feasible set of stops. If S^* is a Pareto solution, then the point

$$(f_{\text{cost}}(S^*), f_{\text{cover}}(S^*))$$

is called an *efficient* point.

To find Pareto solutions we can utilize the following two one-criteria *e*-constraint problems resulting from (BSL).

(BSL-cost(Q)) Given $\mathcal{D}$, $G = (V, E)$ with its set of points $\mathcal{T}$, weights w_d, and norms (or gauges) γ_d for all $d \in \mathcal{D}$, find a set $S^* \subseteq \mathcal{T}$ such that $f_{\text{cover}}(S^*) \geq Q$ and $f_{\text{cost}}(S^*)$ is minimal.

(BSL-cover(k)) Given $\mathcal{D}$, $G = (V, E)$ with its set of points $\mathcal{T}$, weights w_d and norms (or gauges) γ_d for all $d \in \mathcal{D}$, find a set $S^* \subseteq \mathcal{T}$ such that $f_{\text{cost}}(S^*) \leq k$ and $f_{\text{cover}}(S^*)$ is maximal.

Due to Haimes and Chankong [HC83] we have the following result.

Lemma 1 *Let $Q, k \in \mathbb{N}$.*

1. *Let S be a unique optimal solution of (BSL-cost(Q)). Then S is a Pareto solution. If more than one optimal solution of (BSL-cost(Q)) exists, the solutions that additionally maximize f_{cover} are Pareto solutions.*

2. *Let S be a unique optimal solution of (BSL-cover(k)). Then S is a Pareto solution. If more than one optimal solution of (BSL-cover(k)) exists, the solutions that additionally minimize f_{cost} are Pareto solutions.*

Using Lemma 1 to find Pareto solutions is known as the *e-constraint method*, see, e.g., [Ehr00]. Unfortunately, both *e*-constraint problems are hard to solve.

Corollary 1 *(BSL) and the two e-constraint problems (BSL-*cost*) and (BSL-*cover*) are NP-hard, even if all weights w_d are equal to 1.*

Proof: From Theorem 1 ([SHLW02]) we know that (CSL), i.e., finding a minimum cardinality set of stations covering all demand points, is NP-hard. The decision version of both *e*-constraint problems (BSL-cost(Q)) and (BSL-cover(k)) is the following:

Given $\mathcal{D}$, $G = (V, E)$ with its planar embedding $\mathcal{T}$, weights w_d, norms (or gauges) γ_d, does there exist a set $S^* \subseteq \mathcal{T}$ such that $f_{\text{cost}}(S^*) \leq Q$ and $f_{\text{cover}}(S^*) \geq k$?

Defining $Q = \sum_{d \in \mathcal{D}} w_d$ shows that the decision version of (CSL) is a special case of the decision version of both (BSL-cost(Q)) and (BSL-cover(k)) and thus both *e*-constraint problems are NP-hard.

<div align="right">QED</div>

We now discuss the two lexicographic optimal solutions, for which we know that they are Pareto solutions.

- Maximizing f_{cover} as first objective means that we have to cover **all** demand points, that can be covered, i.e., all demand points $d \in \text{cover}(\mathcal{T})$. This yields (CSL) as in [SHLW02], if we define

$$\mathcal{D}' = \mathcal{D} \cap \text{cover}(\mathcal{T})$$

  as the set of demand points to be covered, and hence this problem is NP-hard (see Theorem 1).

- On the other hand, minimizing f_{cost} leads to a trivial problem since it can be solved easily by not installing any stop at all.

Note that the constraint version of (BSL-cover(k)), i.e., to locate at most k stops in such a way that f_{cover} is maximized, was investigated in [KPS$^+$02] for the case of one single straight-line track and for the case of two parallel straight-line tracks. For both cases, polynomial time algorithms using dynamic programming were developed with a time complexity of $O(k|\mathcal{D}|^2)$

for the single track case. Moreover, it is shown that along one straight line track, the unweighted version (BSL-cover) is equivalent to a one-dimensional uncapacitated and unimodular k-facility location problem. As observed by [Tam02] the problem can hence be solved in $O(k|\mathcal{D}|log(\mathcal{D}))$ time.

3 Integer programming formulations

To derive integer programming formulations we use the methodology developed in [SS02]. For an edge $e \in E$ with endpoints v_1^e, v_2^e we define

$$\mathcal{T}^e(d) = \{s \in e : \gamma_d(d, s) \leq r\}$$

as the set of all points on the edge $e \subseteq \mathcal{T}$ that can be used to cover demand point d, and

$$\mathcal{T}(d) = \{s \in \mathcal{T} : \gamma_d(d, s) \leq r\}.$$

Note that $s \in \mathcal{T}(d)$ if and only if $d \in \text{cover}(s)$. The following simple observation will become important later.

Lemma 2 *For each demand point $d \in \mathbb{R}^2$ the set $\mathcal{T}^e(d)$ is an interval contained in edge e.*

Proof: Note that $\mathcal{T}^e(d) = e \cap \{x \in \mathbb{R}^2 : \gamma_d(d, x) \leq r\}$ is the intersection of two convex sets, namely, of the line segment e and the unit ball of the norm (or gauge) γ_d about d. Consequently, $\mathcal{T}^e(d)$ itself is a convex set contained in a line segment and hence a line segment itself.

<div align="right">QED</div>

Let f_d^e, l_d^e denote the endpoints of the interval $\mathcal{T}^e(d)$ (which may coincide with the endpoints v_1^e, v_2^e of the edge e). We write

$$[f_d^e, l_d^e] = \mathcal{T}^e(d).$$

Along the lines of [SHLW02] we can now derive a finite dominating set $\mathcal{S} \subseteq \mathcal{T}$ as follows. For each edge let

$$\mathcal{S}^e = \bigcup_{d \in \mathcal{D}} \{f_d^e, l_d^e\} \cup \{v_1^e, v_2^e\}$$
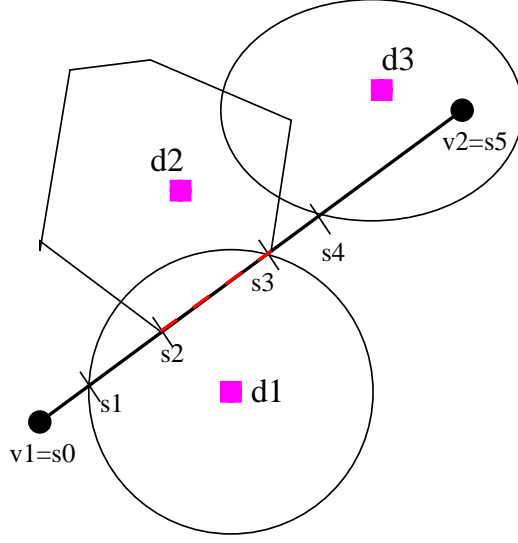
Figure 2: The set of candidates along one linear piece $e \in E$.

be the set of all endpoints of intervals $\mathcal{T}^e(d)$. This set can be ordered along the edge e (e.g., by starting in v_1^e and moving to v_2^e), resulting in a set

$$\mathcal{S}^e = \{s_1, s_2, \ldots, s_{N_e}\},$$

and we write $v_1^e = s_1 < s_2 < \ldots < s_{N_e} = v_2^e$ to indicate the order of the points with respect to $v_1^e < v_2^e$. In the following we show that

$$\mathcal{S} = \bigcup_{e \in E} \mathcal{S}^e$$

is a finite dominating set for the bicriterial stop location problem. For an illustration of $\mathcal{S}$ we refer to Figure 2.

Lemma 3 *Let e be an edge of E, and let $s \in]s_j, s_{j+1}[_e$ for some $j \in \{0, 1, \ldots, N_e\}$. Then*
$$\mathrm{cover}(s) \subseteq \mathrm{cover}(s_j) \cap \mathrm{cover}(s_{j+1}).$$

Proof: Suppose $\mathrm{cover}(s) \not\subseteq \mathrm{cover}(s_j)$, i.e., there exists $d \in \mathcal{D}$ such that $\gamma_d(d, s) \leq r$ and $\gamma_d(d, s_j) > r$. In other words, $s \in \mathcal{T}^e(d)$, but $s_j \notin \mathcal{T}^e(d)$. From Lemma 2 we know that $\mathcal{T}^e(d)$ is an interval. Hence, an endpoint of

8

$\mathcal{T}^e(d)$ lies between s and s_j. This contradicts the definition of s_j and s_{j+1} as consecutive points in $\mathcal{S}^e$.

<div align="right">QED</div>

Theorem 2 $\mathcal{S}$ *is a finite dominating set for (BSL-cost(Q)), (BSL-cover(k)), and for (BSL) for $Q, k \in \mathbb{N}$. More precisely,*

- *Either (BSL-cost(Q)) is infeasible, or it has an optimal solution $S^* \subseteq \mathcal{S}$.*

- *Either (BSL-cover(k)) is infeasible, or it has an optimal solution $S^* \subseteq \mathcal{S}$.*

- *Let (k, Q) be an efficient solution of (BSL). Then there exists a Pareto solution $S \in \mathcal{S}$ with $f_{\text{cost}}(S) = k$ and $f_{\text{cover}}(S) = Q$.*

Proof: Given some optimal (or Pareto) set S^*, we iteratively construct a set $S' \subseteq \mathcal{S}$ by moving stops of the given set S^* into points of $\mathcal{S}$ without changing the objective function values as follows. Let $s \in S^* \setminus \mathcal{S}$ be a point in the optimal solution and let e be the edge of s. Then determine two consecutive points $s_j, s_{j+1} \in \mathcal{S}^e$ such that s lies between s_j and s_{j+1}. According to Lemma 3 we know that $\text{cover}(s) \subseteq \text{cover}(s_j)$, hence

$$S' = S^* \setminus \{s\} \cup \{s_j\}$$

satisfies

$$\begin{aligned} f_{\text{cover}}(S^*) &\leq f_{\text{cover}}(S') \text{ and} \\ f_{\text{cost}}(S^*) &\geq f_{\text{cost}}(S'), \end{aligned}$$

i.e., S' is at least as good as S^* with respect to both criteria. Proceeding like this for all points in $S^* \setminus \mathcal{S}$ proves the result.

<div align="right">QED</div>

Using Theorem 2, (BSL) and its two e-constraint problems can be formulated as integer programs. As decision variable we define

$$x_s = \begin{cases} 1 & \text{if candidate } s \text{ is chosen as a new stop} \\ 0 & \text{otherwise} \end{cases}.$$

<div align="center">9</div>

To keep track of the population covered by the new stops, we also have to know, which demand points are covered and which not. We therefore define another set of binary variables

$$
y_d = \begin{cases} 1 & \text{if demand point } d \text{ is covered} \\ 0 & \text{otherwise} \end{cases} ,
$$

and let $w = (w_{d_1}, w_{d_2}, \ldots, w_{d_{|\mathcal{D}|}})$ and $\mathbb{1} \in \mathbb{R}^{|\mathcal{S}|}$ be the vector with a 1 in each component.

Furthermore, we can store the covering information in the following *covering matrix* $A^{\mathrm{cov}} = (a_{ds})$ with

$$
a_{ds} = \begin{cases} 1 & \text{if } d \in \mathrm{cover}(s) \text{ (or, equivalently, if } s \in \mathcal{T}(d)) \\ 0 & \text{otherwise} \end{cases} ,
$$

The IP model of (BSL) can now be formulated as

$$
\begin{aligned}
\min \quad & \begin{pmatrix} \mathbb{1}x \\ -wy \end{pmatrix} \\
\text{s.t.} \quad A^{\mathrm{cov}} x & \geq y \\
x & \in \{0,1\}^{|\mathcal{S}|} \\
y & \in \{0,1\}^{|\mathcal{D}|}.
\end{aligned}
$$

The IP model for (BSL-cost(Q)) is

$$
\begin{aligned}
\min \quad & \mathbb{1}x \\
\text{s.t.} \quad A^{\mathrm{cov}} x - y & \geq 0 \\
wy & \geq Q \\
x & \in \{0,1\}^{|\mathcal{S}|} \\
y & \in \{0,1\}^{|\mathcal{D}|},
\end{aligned}
$$

beginequation

and (BSL-cover(k)) is given by

$$
\begin{aligned}
\max \quad & wy \\
\text{s.t.} \quad A^{\mathrm{cov}} x - y & \geq 0 \\
\mathbb{1}x & \leq k \\
x & \in \{0,1\}^{|\mathcal{S}|} \\
y & \in \{0,1\}^{|\mathcal{D}|}.
\end{aligned}
$$

4 Bicriterial stop location along a polygonal line

We now analyze the situation along a polygonal line $\mathcal{T}$.

Lemma 4 *If $\mathcal{T}$ is a polygonal line and $\mathcal{T}(d)$ is connected for each demand point d, then A^{cov} has the consecutive ones property, i.e., in each row of A^{cov} the ones appear consecutively.*

Proof: Let $a_{ds_1} = a_{ds_2} = 1$ for $s_1 < s_2$. We then have to show that $a_{ds} = 1$ for all s with $s_1 < s < s_2$. Take a candidate s on the polygonal line between s_1 and s_2. From $a_{ds_1} = a_{ds_2} = 1$ we know that $s_1, s_2 \in \mathcal{T}(d)$. Hence, since $\mathcal{T}(d)$ is connected, also $s \in \mathcal{T}(d)$ and hence $a_{ds} = 1$.

QED

Note that the assumption of Lemma 4 is always satisfied if $\mathcal{T}$ consists of one single edge only, an observation which was first noted in [SHLW02]. Generalizations and decomposition results that can be used to apply this fact to more complex networks are given in [Sch02].

To illustrate the condition of Lemma 4 we consider Figure 3 and Figure 4. In figure 3 an example of a polygonal line not satisfying the condition of Lemma 4 and with a coefficient matrix without consecutive ones property is given. In this example, $\mathcal{T}$ is a polygonal line consisting of three nodes. Numbering the candidates from left to right, A^{cov} is given by

$$A^{\mathrm{cov}} = \begin{pmatrix} 1 & 1 & 1 & 0 & 0 & 0 & 1 & 1 & 0 \\ 0 & 1 & 1 & 1 & 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 & 1 & 1 & 1 & 0 \end{pmatrix},$$

which cannot be reordered to satisfy the consecutive ones property.

On the other hand, Figure 4 shows an example for a polygonal line together with a set of demand points $\mathcal{D}$ satisfying the consecutive ones property.

The importance of Lemma 4 is due to the fact that matrices having the consecutive ones property are totally unimodular such that in this case the stop location problem (CSL) of [SHLW02] can be solved efficiently by linear programming methods. Unfortunately, even if A^{cov} has the consecutive ones property and $w_d = 1$ for all $d \in \mathcal{D}$, this property needs not hold for the
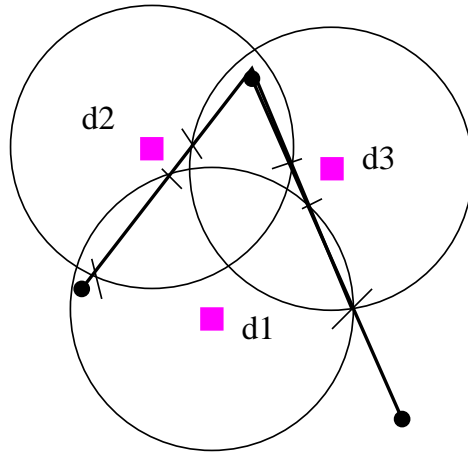
Figure 3: An instance of (BSL) on a polygonal line where $\mathcal{T}(d_1)$ is not connected, and without consecutive ones property.
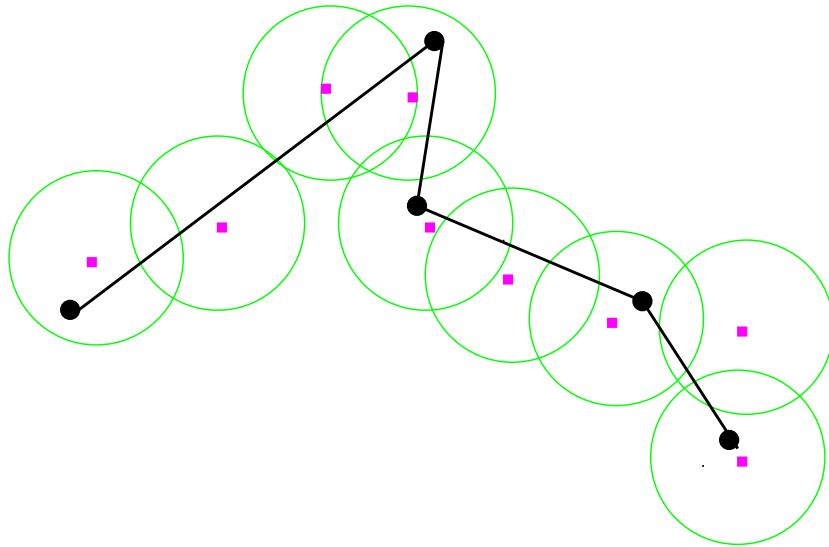

Figure 4: An instance of (BSL) on a polygonal line satisfying that all sets $\mathcal{T}(d)$ are connected, and hence having the consecutive ones property.
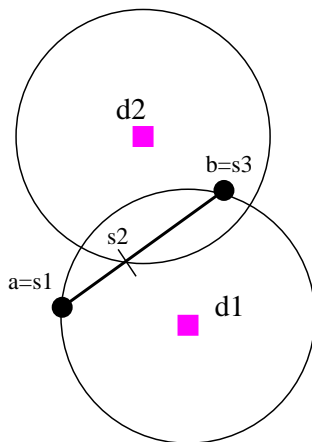
Figure 5: The coefficient matrix of (BSL-cost) is not totally unimodular.

constraint versions of our problem (BSL) as the following example demonstrates.

Consider Figure 5 and note that the coefficient matrix in this small example is

$$A^{\mathrm{cov}} = \begin{pmatrix} 1 & 1 & 1 \\ 0 & 1 & 1 \end{pmatrix},$$

which has the consecutive ones property.

(**BSL-**cost(Q)): Although A^{cov} has the consecutive ones property that does not yield a totally unimodular coefficient matrix for (BSL-cost(Q)). Namely, the coefficient matrix of (BSL-cost) is in this example given as

$$\begin{pmatrix} 1 & 1 & 1 & -1 & 0 \\ 0 & 1 & 1 & 0 & -1 \\ 0 & 0 & 0 & 1 & 1 \end{pmatrix},$$

which is not totally unimodular.

(**BSL-**cover(k)): On the other hand, using the same example for (BSL-cover(k)) the coefficient matrix is given by

$$\begin{pmatrix} 1 & 1 & 1 & -1 & 0 \\ 0 & 1 & 1 & 0 & -1 \\ -1 & -1 & -1 & 0 & 0 \end{pmatrix},$$

13

which does not have the consecutive ones property, but still is a totally unimodular matrix.

This observation holds in general.

Lemma 5 *Let A^{cov} have the consecutive ones property and assume that $w_d = 1$ for all $d \in \mathcal{D}$. Then (BSL-cover(k)) can be solved by linear programming.*

Proof: Note that $\begin{pmatrix} A^{\mathrm{cov}} \\ 1\ 1\ \ldots\ 1 \end{pmatrix}$ has the consecutive ones property and hence is totally unimodular. Thus, also $\begin{pmatrix} A^{\mathrm{cov}} \\ -1\ -1\ \ldots\ -1 \end{pmatrix}$ is totally unimodular and hence the coefficient matrix

$$\begin{pmatrix} A^{\mathrm{cov}} & -I \\ -1\ -1\ \ldots\ -1 & 0\ 0\ \ldots\ 0 \end{pmatrix}$$

of the IP-formulation of (BSL-cover(k)) also satisfies this property. Consequently, the result follows from integer programming theory, see, e.g., [NW88].

<div align="right">QED</div>

This observation (although not true for arbitrary weights w_d) motivates to solve a family of e-constraint problems of type (BSL-cover) to find all efficient solutions of (BSL) in the case that A^{cov} has the consecutive ones property. In the next section we show how this can be done efficiently by dynamic programming.

5   Dynamic programming approach

To develop a dynamic programming approach for (BSL) we first investigate cover(S) in more detail. Again, consider a polygonal line $\mathcal{T}$ satisfying the assumption of Lemma 4 and let $\mathcal{S}$ be the set of candidates. We assume that the candidates have been ordered along $\mathcal{T}$, e.g., from left to right.

Lemma 6 *Let $\mathcal{T}$ be a polygonal line satisfying that $\mathcal{T}(d)$ is connected for all $d \in \mathcal{D}$. Let $S = \{s_1, \ldots, s_p\} \subseteq \mathcal{T}$ with $s_1 < \ldots < s_p$. Then for all $i = 1, \ldots, p-1$ we have*

$$\mathrm{cover}(s_{i+1}) \setminus \mathrm{cover}\{s_1, \ldots, s_i\} = \mathrm{cover}(s_{i+1}) \setminus \mathrm{cover}(s_i).$$

Proof: Since "$\subseteq$" is trivial, we only need to verify "$\supseteq$".

To this end, let $d \in \text{cover}(s_{i+1}) \setminus \text{cover}(s_i)$. We show that $d \notin \text{cover}(s_j)$ for all $j \leq i$. Assume to the contrary that $d \in \text{cover}(s_j)$ for some $j < i$ but that $d \in \text{cover}(s_{i+1})$. This means that $s_j \in \mathcal{T}(d)$ and $s_{i+1} \in \mathcal{T}(d)$, and, since $\mathcal{T}(d)$ is connected due to our assumption also $s_i \in \mathcal{T}(d)$, a contradiction to $d \notin \text{cover}(s_i)$.

<div align="right">QED</div>

Lemma 6 motivates the following dynamic programming approach. In each step we are looking for a set of k stops all of them smaller than (i.e. on the left hand side of) a given stop s_j which itself should be contained in S. This problem is denoted $(\text{P}(k, s_j))$ for some integer k and $s_j \in \mathcal{S}$, and can formally be given as follows.

$(\mathbf{P}(k, s_j))$

$$\min\{f_{\text{cover}}(S) : S \subseteq \{s_1, \ldots, s_j\}, s_j \in S, \text{ and } |S| \leq k\}.$$

Furthermore, define

$$w_{ij} = \sum_{d \in \text{cover}(s_j) \setminus \text{cover}(s_i)} w_d \quad \text{for } i < j \quad \text{and}$$
$$W = f_{\text{cover}}(\mathcal{S}).$$

W denotes the maximum weight which we can cover, if we choose all candidates as new stops, while w_{ij} gives the gain if we add s_j to a set of stops containing s_i as its rightmost stop.

Algorithm :   Finding all efficient solutions of (BSL)

Input: $\mathcal{D}$, a polygonal line $\mathcal{T}$ with connected sets $\mathcal{T}(d)$, weights w.

Output: All efficient solutions of (BSL), and a Pareto solution for each of them.

Step 1. Derive the set of candidates $\mathcal{S} = \{s_1, s_2, \ldots, s_N\}$ as in Theorem 2 and order them along $\mathcal{T}$.

Step 2. Let $W = f_{\text{cover}}(\mathcal{S})$ and $w_{ij} = \sum_{d \in \text{cover}(s_j) \setminus \text{cover}(s_i)} w_d$ for all $i < j$ with $i, j \in \{1, \ldots, N\}$

<div align="center">15</div>

Step 3. Let for all $j = 1, \ldots, N$: $h^1(s_j) = \text{cover}(s_j)$, $S^1(s_j) = \{s_j\}$.

Step 4. For all $j = 1, \ldots, N$:

$$h^k(s_j) = \max\{\max_{i:s_i < s_j} w_{ij} + h^{k-1}(s_i), h^{k-1}(s_j)\}$$

- If $h^k(s_j) = w_{i^*j} + h^{k-1}(s_{i^*})$ let $S^k(s_j) = S^{k-1}(s_{i^*}) \cup \{s_j\}$.
- If $h^k(s_j) = h^{k-1}(s_j)$ let $S^k(s_j) = S^{k-1}(s_j)$.

Step 5. Let $h^k = \max_{j=1,\ldots,N} h^k(s_j) =: h^k(s^*)$ and let $S^k = s^k(s^*)$.

- If $h^k = W$ then set $K^* = k$ and stop.
- Otherwise $k = k + 1$ and goto step 4.

Step 6. Output: $Eff = \{(h^k, k) : k = 1, \ldots, K^*\}$ with corresponding Pareto solutions S^k, $k = 1, \ldots, K^*$.

To show the correctness of the algorithm we need the following lemmas.

Lemma 7 $S^k(s_j)$ *is an optimal solution of $(P(k, s_j))$ with objective value* $h^k(s_j)$.

Proof: We use induction over k. For $k = 1$ the optimal solution of $(P(1, s_j))$ is $S^1(s_j) = \{s_j\}$. Now assume that $S^{k-1}(s_j)$ is the optimal solution of $(P(k-1, s_j))$ for any fixed s_j. For the induction step we first note that $(P(k, s_j))$ is equivalent to

$$\min\{f_{\text{cover}}(S' \cup \{s_j\}) : S' \subseteq \{s_1, \ldots, s_{j-1}\}, \text{ and } |S'| \leq k - 1\}.$$

Now calculate that for any $S' = \{s_{i_1}, s_{i_2}, \ldots, s_{i_p}\}$ with $s_{i_1} < s_{i_2} < \ldots s_{i_p} < s_j$ we obtain

$$
\begin{aligned}
f_{\text{cover}}(S' \cup \{s_j\}) &= \sum_{d \in \text{cover}(S' \cup \{s_j\})} w_d \\
&= f_{\text{cover}}(S') + \sum_{d \in \text{cover}(s_j) \backslash \text{cover}(S')} w_d \\
&= f_{\text{cover}}(S') + \sum_{d \in \text{cover}(s_j) \backslash \text{cover}(s_{i_p})} w_d \\
&\quad \text{due to Lemma 6} \\
&= f_{\text{cover}}(S') + w_{i_p j}
\end{aligned}
$$

16

Hence, $(P(k, s_j))$ can further be rewritten to

$$\min\{f_{\text{cover}}(S') + w_{i_p j} \quad : \quad i_p \in \{s_1, \ldots, s_{j-1}\}, S' \subseteq \{s_1, \ldots, s_{i_p}\},$$
$$i_p \in S', \text{ and } |S'| \leq k - 1\}$$

and it becomes clear that the set S' in this formulation is an optimal solution of $(P(k - 1, s_{i_p}))$. Using the induction hypothesis we finally obtain that $(P(k, s_j))$ is equivalent to

$$\min\{f_{\text{cover}}(S^{k-1}(i_p) + w_{i_p j} : i_p \in \{s_1, \ldots, s_{j-1}\}\}$$

which shows the result.

<div align="right">QED</div>

Corollary 2 S^k *is an optimal solution of (BSL-cover(k)) and its objective value is* h^k.

Proof: This consequence follows from Lemma 7 and the definition of S^k in step 5 of the algorithm.

<div align="right">QED</div>

Finally, to apply Lemma 1 we need the following result.

Lemma 8 *For* $k \leq K^*$ *any optimal solution* S^* *of (BSL-cover(k)) satisfies* $|S^*| = k$.

Proof: Let S be an optimal solution of (BSL-cover(k)) for some $k < K^*$. This means that $f_{\text{cover}}(S) = f_{\text{cover}}(S^k) < W$ due to Corollary 2 and step 6 of the algorithm. Hence there exists $s \notin S$ such that

$$\sum_{d \in \text{cover}(s) \backslash \text{cover}(S)} w_d > 0$$

and hence $f_{\text{cover}}(S \cup \{s\}) > f_{\text{cover}}(S)$. If $|S| \leq k - 1$ this yields that $S \cup \{s\}$ does not contain more than k stops and hence is feasible for (BSL-cover(k)), which is a contradiction to the optimality of S.

<div align="right">QED</div>

Theorem 3 *The algorithm finds all efficient solutions of (BSL).*

Proof: For each $k \leq K^*$ we know from Corollary 2 that S^k is an optimal solution of (BSL-cover(k)) . Furthermore, Lemma 8 shows that all optimal solutions of (BSL-cover(k)) consist of the same number k of stops. Hence (h^k, k) is an efficient solution according to Lemma 1.

On the other hand, no solution S with $|S| > K^*$ is Pareto, since such a solution S always is dominated by S^{K^*}, using that

$$
\begin{aligned}
|S| &> K^* = |S^{K^*}| \text{ and} \\
f_{\text{cover}}(S) &\leq W = f_{\text{cover}}(S^{K^*}).
\end{aligned}
$$

QED

Note that the above algorithm is an application of the algorithm of Bellman-Ford (see, [Bel58, FF62]) to the underlying set covering problem, see [Sch03].

Since the number of candidates $|\mathcal{S}|$ is at most twice the number of demand points for a polygonal line satisfying the assumptions of the algorithm, the worst-case complexity of the algorithm (finding all efficient solutions of the bicriterial stop location problem) is given by $O(K^*|\mathcal{D}|^2)$ where K^* is the minimum number of stops needed to cover all demand points in cover($\mathcal{T}$).

6   Conclusion

In this paper we developed a model for the bicriterial stop location problem and proposed an efficient solution approach for determining a Pareto solution for each efficient solution in the special case that the set of tracks is given by a polygonal line with connected intervals $\mathcal{T}(d)$ for each demand point d. Investigating the real-world data of German rail (DB) all over Germany, it turns out that this assumption is *almost* satisfied in practice. Dealing with a few demand points not satisfying the connectedness of $\mathcal{T}(d)$ (or, more specific, with a few rows of the covering matrix not satisfying the consecutive ones property) is currently under research ([RS03]).

Moreover, the extension of the results to demand regions instead of demand points is investigated. For some first result in this area we refer to [SS02].

References

[Bel58]   R. Bellman. On a routing problem. *Quarterly Applied Mathematics*, 16:87–90, 1958.

[CFT99]   A. Caprara, M. Fischetti, and P. Toth. A heuristic method for the set covering problem. *Operations Research*, 47(5):730–743, 1999.

[DAL82]   M. J. Demetsky, M. Asce, and B. B.-M. Lin. Bus stop location and design. *Transportation Engineering Journal*, 108:313–327, 1982.

[Ehr00]   M. Ehrgott. *Multiple Criteria Optimization*, volume 491 of *Lecture Notes in Economics and Mathematical Systems*. Springer, Berlin, 2000.

[FF62]   L. R. Ford and D. R. Fulkerson. *Flows in Networks*. Princeton University Press, 1962.

[HC83]   Y.Y. Haimes and V. Chankong. *Multiobjective Decision Making — Theory and Methodology*. North Holland, New York, 1983.

[HLS+01]   H.W. Hamacher, A. Liebers, A. Schöbel, D. Wagner, and F. Wagner. Locating new stops in a railway network. *Electronic Notes in Theoretical Computer Science*, 50(1), 2001.

[KPS+02]   E. Kranakis, P. Penna, K. Schlude, D.S. Taylor, and P. Widmayer. Improving customer proximity to railway stations. Technical report, ETH Zürich, 2002.

[LMO02]   G. Laporte, J.A. Mesa, and F.A. Ortega. Locating stations on rapid transit lines. *Computers and Operations Research*, 29:741–759, 2002.

[MDSF98] A. Murray, R. Davis, R.J. Stimson, and L. Ferreira. Public transportation access. *Transportation Research D*, 3(5):319–328, 1998.

[Min67]   H. Minkowski. *Gesammelte Abhandlungen, Band 2*. Chelsea Publishing Company, New York, 1967.

[Mur01a]   A. Murray. Coverage models for improving public transit system accessibility and expanding access. Technical report, Department of Geography, Ohio State University, 2001.

[Mur01b]   A. Murray. Strategic analysis of public transport coverage. *Socio-Economic Planning Sciences*, 35:175–188, 2001.

[NW88]   G. L. Nemhauser and L. A. Wolsey. *Integer and Combinatorial Optimization*. Wiley, 1988.

[RS03]   N. Ruf and A. Schöbel. Set covering problems with almost consecutive ones property. Technical report, Universität Kaiserslautern, 2003. working paper.

[Sch02]   A. Schöbel. Customer-oriented optimization in public transportation, 2002. Habilitationsschrift, Universität Kaiserslautern.

[Sch03]   A. Schöbel. Set covering problems with consecutive ones property. Technical report, Universität Kaiserslautern, 2003. working paper.

[SHLW02] A. Schöbel, H.W. Hamacher, A. Liebers, and D. Wagner. The continuous stop location problem in public transportation. Technical report, Universität Kaiserslautern, 2002. Report in Wirtschaftsmathematik Nr. 81/2001.

[SS02]   A. Schöbel and M. Schröder. Covering population areas by railway stops. In *Proceedings of OR 2002, Klagenfurt*, 2002. to appear.

[Ste89]   R.E. Steuer. *Multiple Criteria Optimization: Theory, Computation, and Application*. Krieger Publishing Company, Malabar, Florida, 1989.

[Tam02]   A. Tamir, July 2002. personal communication.

[TSRB71] C. Toregas, R. Swain, C. ReVelle, and L. Bergman. The location of emergency facilities. *Operations Research*, 19:1363–1373, 1971.