Locating New Stops in a Railway Network [⋆]

Horst W. Hamacher [a,1], Annegret Liebers [b,2], Anita Schöbel [a,3], Dorothea Wagner [b,4], Frank Wagner [c,5]

[a] *University of Kaiserslautern, Department of Mathematics, and Fraunhofer Institute ITWM, 67653 Kaiserslautern, Germany*

[b] *University of Konstanz, Department of Computer and Information Science, 78457 Konstanz, Germany*

[c] *Deutsche Bahn AG, Konzernentwicklung GSV, Potsdamer Platz 2, 10785 Berlin, Germany*

Abstract

Given a railway network together with information on the population and their use of the railway infrastructure, we are considering the effects of introducing new train stops in the existing railway network. One effect concerns the accessibility of the railway infrastructure to the population, measured in how far people live from their nearest train stop. The second effect we study is the change in travel time for the railway customers that is induced by new train stops. Based on these two models, we introduce two combinatorial optimization problems and give NP-hardness results for them. We suggest an algorithmic approach for the model based on travel time and present a real-world application with its first experimental results.

1 Introduction

The motivation for the problems studied in this paper is to increase the attractiveness of train travel in an existing railway network for local traffic, for example to sports and leisure activities or for the daily commuter trip to work, in not too costly a manner. So assume we are given an existing railway network as well as information on people (potentially) using this network. A relatively cheap way of increasing the attractiveness of train travel is then to open up

new train stops along the already existing railway lines, where a new train stop is meant to be a small stop that is cheap to build, and that will typically be served only by the local commuter trains.

We will consider two effects of such new train stops. The first one concerns the accessibility of the railway network: Assuming that living closer to a train stop increases the attractiveness of the railway network (and ultimately increases the tendency to use a train instead of other means of transport), we study the problem of locating (few) new train stops so that (many) people live within a short distance from a train stop. This *accessibility model* will be specified more precisely in Section 2, and a corresponding optimization problem will be shown to be NP-complete.

The second effect of new train stops we will study concerns the travel time for the customers already using the railway network. On the one hand, passengers sitting in a train that now stops at the newly opened train stops are slowed down by the additional stops. On the other hand, customers living closer to a newly opened train stop than to any previously existing train stop need less time to reach their closest train stop. Summing up each of these two contrary effects over all railway customers and subtracting the sums, we obtain a measurement for the effect in overall travel time that a set of new train stops is having for the existing railway customers. Section 3 will specify this *travel time model* and also give an NP-completeness result for the corresponding optimization problem. Section 4 will suggest an algorithmic approach for this optimization problem based on a genetic algorithm, and Section 5 will give experimental results for the complete railway network of the *Deutsche Bahn AG* in Germany.

2 The Accessibility Model

Let $\mathcal{P}$ denote the set of coordinate points, each of them representing a settlement, and let $\mathcal{L}$ be the set of feasible points along the tracks of the railway company. We assume that the set of tracks is piecewise linear, i.e. it is given as the union of a finite number of line segments $S_i \in \mathbb{R}^2$, $i = 1, \ldots, L$:

$$\mathcal{L} = \{x \in \mathbb{R}^2 : x \in S_i \text{ for some } i = 1, \ldots, L\}.$$

A solution X to the problem is given by a set of points in the plane, representing the stops that should be installed. X is feasible, if $X \subseteq \mathcal{L}$, i.e., if each $x \in X$ satisfies $x \in \mathcal{L}$.

To evaluate a feasible solution X, we are interested in how many settlements are covered by X. A settlement $p \in \mathcal{P}$ is *covered* by a point $x \in \mathcal{L}$, if

$$l_2(p, x) \leq r,$$

where l_2 refers to the Euclidean distance (or any other norm), and r is some

given radius. A settlement p is covered by a set $X \subseteq \mathcal{L}$ if

$$l_2(p, x) \leq r \text{ for some } x \in X.$$

For a set X we define

$$\text{cover}(X) := \{p \in \mathcal{P} : l_2(p, X) \leq r\},$$

where $l_2(p, X) := \min_{x \in X} l_2(x, p)$ for $|X| \geq 1$ and $l_2(p, \emptyset) := \infty$. Figures 1 and 2 illustrate how different sets of train stops cover different sets of settlements, using a radius of $r = 2$ km.

Furthermore, we are interested in the number of new stops: We would like to cover as many settlements with as few stops as possible. This goal leads to the definition of the following optimization problem, which unfortunately is NP-complete:

Definition 2.1 [Covering Along Lines] Given a set $\mathcal{P}$ of integer-coordinate points in the plane, a connected set $\mathcal{L}$ which is given as the union of a finite number of line segments in the plane, and positive integers d and $K < |\mathcal{P}|$, can the points of $\mathcal{P}$ be covered by at most K discs of diameter d, all with center points in $\mathcal{L}$?

Theorem 2.2 *Covering Along Lines is NP-complete.*

Proof. We reduce Geometric Covering by Discs to the above problem. Geometric Covering by Discs has been shown to be NP-complete [3] and can be stated as follows: Given a set $\mathcal{P}$ of integer-coordinate points in the plane and positive integers d and $K < |\mathcal{P}|$, can the points of $\mathcal{P}$ be covered by at most K discs of diameter d?

Given an instance of Geometric Covering by Discs, use the same set $\mathcal{P}$ of points and the same numbers d and K and define the set $\mathcal{L}$ as follows to construct an instance of Covering Along Lines: For each unordered pair of points p_1 and p_2 from $\mathcal{P}$,

 (i) add the points forming the line segment between p_1 and p_2 to $\mathcal{L}$ and

(ii) add the points forming a sufficiently large piece of the bisector of p_1 and p_2 to $\mathcal{L}$.

Claim: $\mathcal{P}$ can be covered by at most K discs of diameter d, if and only if $\mathcal{P}$ can be covered by at most K discs of diameter d which all have their center points in $\mathcal{L}$.

To see this, first assume that $\mathcal{P}$ can be covered by some collection C, consisting of at most K discs of diameter d. Then, for each disc $c \in C$:

• If c contains no points of $\mathcal{P}$, then disregard c.

• If c contains only one point p of $\mathcal{P}$, then replace c by the disc with center point p (p is in $\mathcal{L}$, since all line segments from p to any other point are in $\mathcal{L}$) and diameter d.

3

- If c contains a set of points A of $\mathcal{P}$ with $|A| \geq 2$ (i.e., c contains more than one point of $\mathcal{P}$), then replace c by a disc with center point q and diameter d, where q is the center point of the smallest enclosing circle of A.

  Since c covers A, the diameter of the smallest enclosing circle is smaller or equal to d, and hence the disc with radius d and center point q also covers A. Note that finding q is a well-known problem of location theory and can be done in linear time [4]. Moreover it is known that q always lies on at least one bisector of points in A [2], such that q satisfies $q \in \mathcal{L}$.

Together, $\mathcal{P}$ can be covered by at most K discs of diameter d, all with center points in $\mathcal{L}$.

The other direction is immediate. □

Note that in the special case of only one line segment in $\mathcal{L}$, Covering Along Lines turns out to be a set covering problem whose constraint matrix is an interval matrix. Hence, in this case the problem is polynomially solvable by linear programming. This result can be extended to sets of line segments, if no settlement can be covered by two stops from different line segments.

3   The Travel Time Model

Some more notation is needed to state the travel time model. We suppose that $\mathcal{L}$ is given as a network $G = (V, E)$ embedded in the plane. Note that all the already existing train stops are contained in V. For a point $x \in \mathcal{L}$ let $e(x) \in E$ denote the edge of G on which x is located. Furthermore, for each edge e the number of customers traveling along e is denoted by w_e. Then for a set X of new stops,

$$t_d(X) = \sum_{x \in X} s \cdot w_{e(x)}$$

gives the amount of additional travel time for the customers, which is caused by the additional stopping activities of the trains. (We assume that a constant time delay s is caused by any additional stop of a train.)

On the other hand, some of the customers will save travel time, since a new train stop may reduce the distance to their closest train stop and hence the time they need to get on a train. In our model, the reduction of the distance for a point $p \in \mathcal{P}$ is calculated by

$$l_2(p, S) - l_2(p, S \cup X),$$

where $S \subseteq V$ denotes the set of already existing train stops, and $l_2(p, Y)$ is the closest Euclidean distance from p to any point in Y as it has been used in the accessibility model before.

To transform the possible reduction of distance into an amount of saved access time we introduce a piecewise linear function $g : \mathbb{R} \times \mathbb{R} \to \mathbb{R}$ in two variables, assigning an amount of saved time to each reduction of the distance,

given as a pair consisting of old and new distance of a settlement from the nearest train stop. For example, g can be defined as

$$g(x, y) := (x - y)/5,$$

assuming an average speed of 5 km/h, or as

$$g(x, y) = \begin{cases} \frac{x-y}{4} & \text{if } x \leq 1 \text{ (customer walks)} \\ \frac{x-y}{7} & \text{if } 1 < x \leq 5 \text{ (customer uses a bike)} \\ \frac{x-y}{20} & \text{if } 5 < x \text{ (customer uses a bus or a car)} \end{cases}$$

Note that this definition would assume that the customer stays with the means of transport used for distance x.

Denoting for each $p \in \mathcal{P}$ the number of customers from the corresponding settlement by v_p, the positive effect to the travel time through saved access time can be calculated by

$$t_a(X) = \sum_{p \in \mathcal{P}} v_p \cdot g(l_2(p, S), l_2(p, S \cup X)).$$

The travel time model can now be summarized:

$$\max t_a(X) - t_d(X)$$

such that

$$x \in \mathcal{L} \text{ for all } x \in X.$$

Note that we neglect the change in train riding time that is caused by starting or ending the trip at a different train stop, assuming that these gains and losses in train riding time roughly even out.

We can now define the corresponding problem Saved Travel Time and show its NP-completeness:

Definition 3.1 [Saved Travel Time] Given a set $\mathcal{P}$ of integer-coordinate points in the plane, each of them with a nonnegative weight v_p, an embedded network $G = (V, E)$ in the plane, each of its arcs with a nonnegative weight w_e, a subset $S \subseteq V$, a positive number s and a real number K, and given a piecewise linear function in two variables $g : \mathbb{R} \times \mathbb{R} \to \mathbb{R}$, is there a set of points X on the edges of G such that

$$f(X) := \sum_{p \in \mathcal{P}} v_p \cdot g(l_2(p, S), l_2(p, S \cup X)) - \sum_{x \in X} s \cdot w_{e(x)} \geq K?$$

Theorem 3.2 *Saved Travel Time is NP-complete.*

Proof. We reduce Covering Along Lines to Saved Travel Time. Given an instance of Covering Along Lines with $\mathcal{P}$, $\mathcal{L}$, d and $K < |\mathcal{P}|$ as in Definition 2.1, construct a network $G = (V, E)$ out of the line segments in $\mathcal{L}$ in the natural

way (i.e. with V being the end points and the intersection points of the line segments, and with E being the parts of the line segments obtained by subdividing them with their intersection points), use the same set of points $\mathcal{P}$, and define

$$r := d/2,$$
$$S := \emptyset,$$
$$v_p := M \text{ for all } p \in \mathcal{P} \text{ with } M > |\mathcal{P}|,$$
$$w_e := 1 \text{ for all } e \in E, \text{ and}$$
$$s := 1.$$

Furthermore, define

$$g(x, y) := \begin{cases} 1 & \text{if } y \leq r \\ 0 & \text{otherwise} \end{cases}$$

With these definitions, we get that

$$f(X) = M \cdot |\mathrm{cover}(X)| - |X|$$

where the cover is defined with respect to the set $\mathcal{P}$.

Claim: There exists a solution X to Saved Travel Time with $f(X) \geq M|\mathcal{P}| - K$ if and only if there exists a solution to Covering Along Lines with no more than K discs.

To see this, assume first that X is a solution to Saved Travel Time with $f(X) \geq M|\mathcal{P}| - K$. Then $\mathrm{cover}(X) = \mathcal{P}$, for otherwise $|\mathrm{cover}(X)| \leq |\mathcal{P}| - 1$ and hence

$$\begin{aligned} f(X) &= M \cdot |\mathrm{cover}(X)| - |X| \\ &\leq M \cdot |\mathrm{cover}(X)| \\ &\leq M|\mathcal{P}| - M \\ &< M|\mathcal{P}| - |\mathcal{P}| \\ &< M|\mathcal{P}| - K \\ &\leq f(X), \end{aligned}$$

a contradiction. Furthermore, $\mathrm{cover}(X) = \mathcal{P}$ implies $|X| \leq K$.

For the other direction, let X be a solution to Covering Along Lines with $\mathrm{cover}(X) = \mathcal{P}$ and $|X| \leq K$. Hence,

$$\begin{aligned} f(X) &= M \cdot |\mathrm{cover}(X)| - |X| \\ &= M|\mathcal{P}| - |X| \\ &\geq M|\mathcal{P}| - K. \end{aligned}$$

$\square$

The travel time model can be viewed as a location problem in which the new facilities must be located on a given network, while the existing facilities are given as points in the Euclidean plane. (For a recent survey of theory and applications in the field of location theory, see [1].) This means that neither planar location models nor network location theory really fits. Moreover, even in the case of locating only one new stop, the objective function of Saved Travel Time is not convex. For both reasons, the classical approaches of location theory are not applicable, and new solution algorithms have to be developed. Our first approach — a genetic algorithm — is presented in the next section.

4 Algorithmic Approach

We propose to use a set of discrete candidates for new train stops, and to then employ a genetic algorithm [5] for the maximization problem Saved Travel Time. Consider a simple set of candidates as illustrated on the left side of Fig. 2: Place four evenly spaced candidates on each line segment of the railroad network, and place a candidate at each end of such a line segment that is not an existing train stop. Clearly, if one of these candidates, c, does not by itself yield a positive saved travel time, then the saved travel time of a set X of new train stops containing c does not decrease if we drop c from X. So such a candidate c is never useful for a solution X, and we reduce the set of candidates by dropping every candidate c for which $t_d(\{c\}) > t_a(\{c\})$. See the right side of Fig. 2 for an illustration. With this reduced set of candidates C for the travel time model, the natural application of a genetic algorithm works with bit vectors of length $|C|$ where each bit stands for exactly one candidate. So $11\ldots1$ describes the whole set C, $00\ldots0$ describes the empty set, and every other bit vector of length $|C|$ describes a set of new train stops consisting of some members of C.

To generate a starting population of i individuals for a genetic algorithm, we either use a fixed probability p to determine for each of the $i \cdot |C|$ bits whether they are set to 1, or, for each bit b, we first determine a number $p_b \in \{0.001, 0.002, \ldots, 0.999\}$ randomly, and then use p_b as the probability to determine whether b is set to 1.

In order to generate the next generation from a given population, we divide the given population in half, where one half contains the individuals with the highest saved travel time. Then among these "better" individuals, we repeatedly choose two different individuals at random to "mate" them. We mate two individuals by first randomly choosing a crossover point $j \in \{1, 2, \ldots, |C|-1\}$ and by then applying the crossover by swapping the tails consisting of $|C| - j$ bits. For example, if two individuals 101100100 and *110110010* with $|C| = 9$ are mated with crossover point $j = 4$, then the two resulting individuals are 1011*10010* and *1101*00100. Finally, we flip some of the bits of the new individuals with a very low probability p_m ("mutation"). So the next generation consists of the better half of the old generation, plus the results of mating
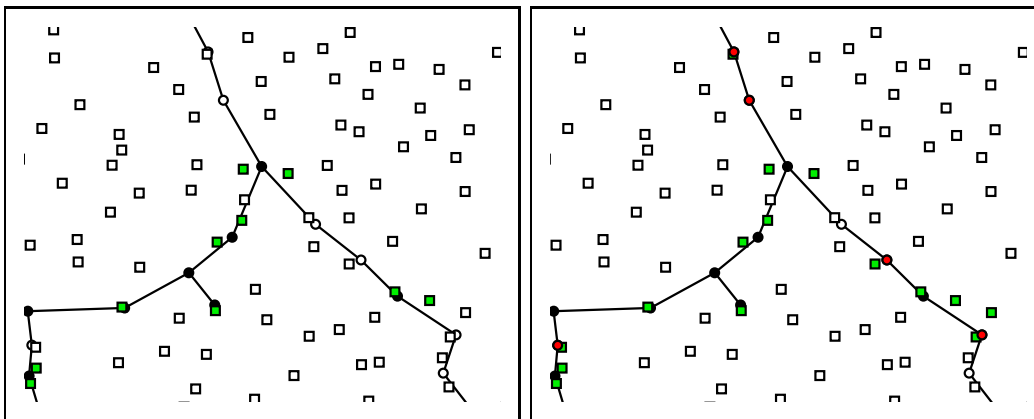
7

Fig. 1. *Left:*The situation with the existing train stops. *Right:* The solution found after 1000 generations as shown in Fig. 3. Black circles depict existing train stops, red (shaded) circles depict new train stops, and green (shaded) squares depict covered settlements.
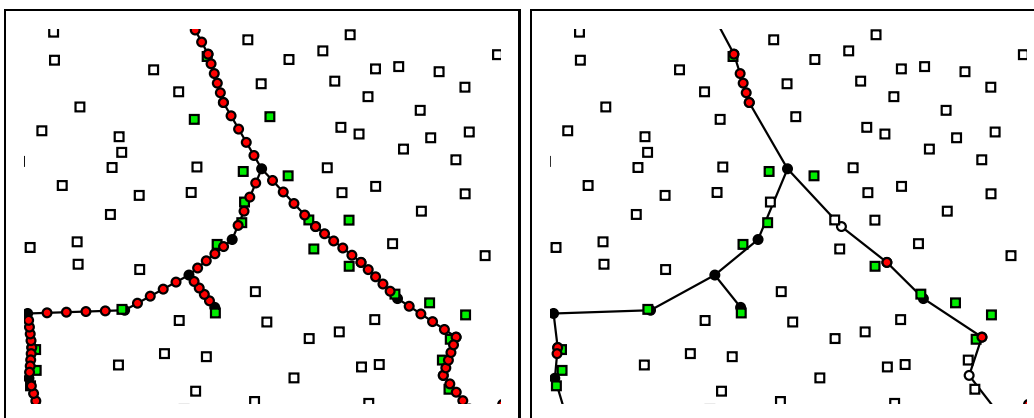

Fig. 2. *Left:*The raw set of candidates. *Right:* The reduced set of candidates.

individuals from this better half with each other as just described.

5   Application and Experimental Results

We applied the genetic algorithm described in Section 4 to a preliminary set of input data describing the situation in Germany. There are 6 900 existing stops with polygonal edges consisting of 8 700 line segments, and there are 30 600 settlements with their population and their estimated number of railway customers. This status quo is depicted for a small excerpt in Southern Germany on the left side of Fig. 1. The set of candidates used for the genetic algorithm consists of 6 700 potential new stops. The raw set of candidates and the reduced set of candidates are illustrated in Fig. 2. Assuming a time delay s of 2 minutes per additional stop and the function $g(x, y) := (x - y)/5$ to describe the gain in access time when the distance to the nearest train stop is reduced from x to y km, we started the genetic algorithm with this

set of candidates and with fixed probabilities $p \in \{0.25, 0.5, 0.75\}$ three times each for a population size of 20. The probability p_m for mutation of bits after crossover was set to 0.0001. After 100 generations each, we let the resulting population with the highest saved travel time so far evolve for another 900 generations. The development of this population over the course of its 1000 generations is shown in Fig. 3. For comparison, we started the genetic algorithm with varying initial probabilities p_b for each bit b of each individual with population size 100, and it turned out that the genetic algorithm finds the right "density" of the solutions very quickly, i.e. we actually do not need to first determine a good probability p for the creation of the starting population: While the starting population with varying initial probabilities p_b contained individuals with almost no candidates as well as individuals with almost all of the 6 700 candidates, the difference between the lowest and the highest number of candidates in the individuals of the 10th generation had already shrunk to less than 1 000, and after 100 generations this difference was a mere 22.

The solution with the highest saved travel time found so far by our genetic approach is illustrated on the right side of Fig. 1. It shows several candidates from the right side of Fig. 2 actually becoming new stops, and in this excerpt, the placement of new train stops looks reasonable and the development of this solution as depicted in Fig. 3 is promising. But when inspecting this solution more closely in other parts of Germany, we noticed that there are in some parts unrealistically many new stops in the solution. It turns out that such effects are caused by inaccuracies in the input data: For example, unrealistically low numbers of customers traveling along edges make placing new stops along these edges too "cheap", resulting in too many new train stops.

These and other problems mean that the solutions found by the genetic algorithm are not to be taken as final suggestions for placing new train stops in order to save travel time, and the amounts of saved travel time indicated on the vertical axis of Fig. 3 have to be read with caution. Rather, we have presented an algorithmic approach, that, if current problems with the input data are overcome, can yield useful results for the placement of new train stops when the goal is to maximize the saved travel time.

6   Conclusion

Studying the effects of introducing new train stops in existing railway networks, we have considered two models, one based on the closeness of the train stops to the population, and the other based on maximizing the overall saved travel time by all customers. For both models, we have shown the NP-hardness of a naturally arising optimization problem. For the travel time model, we have applied a genetic algorithm and we have shown that it can yield useful solutions despite the NP-hardness results. For the first model, however, an algorithmic approach still needs to be developed and applied. Here, the goal
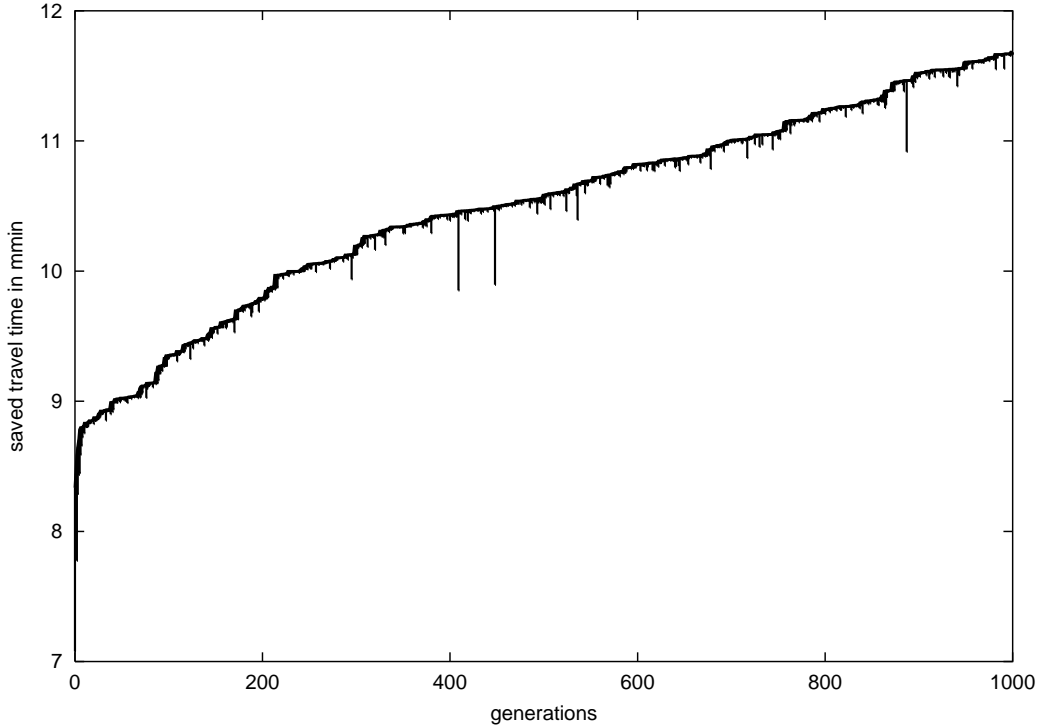
Fig. 3. The development of the objective function *saved travel time* over 1000 generations with population size 20.

would be to find pareto-optimal solutions for the bicriterial problem with objective functions the percentage of population covered by a set of new stops (to be maximized) and the number of new stops (to be minimized).

Furthermore, for both models, a refinement of the effects under consideration is also desirable: Instead of a settlement being either covered or not depending on its distance from a train stop being at most or more than a given radius r, one could consider a settlement being partially covered with the percentage of its covering depending in some more refined way on its distance from a train stop. Similarly, a more realistic function for converting a decrease in distance from a train stop into an amount of saved access time than simply assuming an average speed of 5 km/h is needed.

Finally, other effects of new train stops not considered in our two models include building and operating costs. And it would be of interest to consider a dynamic model where a change in the numbers of customers due to new train stops is also taken into account.

Acknowledgement

References

[1] Drezner, Z. and H. Hamacher, editors, "Location Theory: Applications and Theory," Springer, 2001.

[2] Elzinga, J. and D. Hearn, *Geometrical solutions for some minimax location problems*, Transportation Science **6** (1972), pp. 379–394.

[3] Johnson, D. S., *The NP-Completeness Column: An Ongoing Guide*, J. Algorithms **3** (1982), pp. 182–195.

[4] Megiddo, N., *Linear time algorithms for linear programming in* $\mathbb{R}^3$ *and related problems*, SIAM J. on Computing **12** (1983), pp. 759–776.

[5] Reeves, C. R., *Genetic Algorithms*, in: C. R. Reeves, editor, *Modern Heuristic Techniques for Combinatorial Problems*, McGraw Hill, 1995 pp. 151–196.