Towards transfer synchronization of regularity-based bus operations with sequential hill-climbing

Konstantinos Gkiotsalitis¹ · Nitin Maslekar¹

Accepted: 31 May 2018 / Published online: 30 July 2018
© The Author(s) 2018

Abstract In this work we model and discuss how we can achieve coordination between different bus service lines. Key problem challenges are (a) the multiple conflicting priorities (on one hand the improvement of bus service regularity and on the other hand the reduction of passenger transfer waiting times) and (b) the computational complexity for re-scheduling the dispatching times of bus trips for meeting the conflicting priorities. Initially, a model for reducing the waiting times at bus transfer stations while also improving the operations of regularity-based bus services subject to operational constraints is introduced. Conflicting priorities are handled with the introduction of weight factors that allow bus operators to decide the trade-off between improvement of regularity-based operations and reduction of passenger waiting times at transfer stations. After that, an exterior point penalty function is introduced for handling operational constraints and a sequential hill-climbing search strategy is applied for converging to an approximate optimal solution. For our case study, we utilize general transit feed specification data from two regularity-based bus services in central Stockholm that intersect in five transfer stations. Experimental tests showcase a 13% potential waiting time improvement at transfer stations while sacrificing only 2.8% of service regularity and satisfying all operational constraints.

Keywords Excess waiting time · Multi-line coordination · Bus transfer waiting time · Nonlinear programming

✉ Konstantinos Gkiotsalitis
k.gkiotsalitis@utwente.nl

¹ University of Twente, Faculty of Engineering Technology, Horst - Ring, 7500 AE, Enschede, The Netherlands

1 Introduction

Attempts to improve the service quality of high and low frequency bus services have led to the introduction of several key performance indicators (KPIs) for bus operations such as the On Time Adherence (OTA) to the planned schedule or the excess waiting time (EWT) of passengers at bus stations. OTA is a widely used KPI for punctuality-based bus services where the performance of the bus operations is measured by comparing the real arrival times of buses at bus stops against the planned ones. On the other hand, high-frequency services in metropolitan areas have difficulties adhering to the scheduled arrival times at stops and for this reason the bus operations are run on a regularity basis. Regularity-based bus operations, which are the main focus of this work, do not try to adhere to the scheduled arrival times at the bus stops but aim to achieve a certain level of regularity by avoiding bus bunching and significant headway variations. For instance, each bus trip does not have a specific planned arrival time at the bus stops and the only objective is the minimization of the headway variations among buses (keeping the actual headways as close to the scheduled ones as possible). For this reason, regularity-based services utilize different KPIs such as the EWT which represents the variation of the actual waiting times of passengers at bus stops from the planned ones.

Offline and real-time control measures such as departure time changes, speed control and bus holding have been proposed from several studies that try to meet the Public Transport Authorities (PTA) KPIs related to passenger waiting times at stops and in-vehicle travel times (Dessouky et al. 2003; Cats et al. 2012; Gkiotsalitis and Maslekar 2015; Fu and Yang 2002; Xuan et al. 2011). Other works, such as Yu et al. (2011), Chowdhury and Chien (2011), were more general by not focusing on PTA KPIs but trying to find an acceptable balance between passenger and operational costs to maximize service quality while reducing operational costs. Apart from meeting the PTA KPIs for each individual bus service, bus operators should also reduce the total travel time of passengers to ensure that their services remain competitive. This was also discussed in Botea et al. (2013), Botea and Braghin (2015) that focused on the significant travel time increase of passengers who change several modes due to the operational uncertainty.

Reducing the overall passenger travel time for passengers that use multiple lines for completing their journeys requires the minimization of the waiting times at bus stations that serve as interchange points among different bus services. Meeting the bus service KPIs while ensuring acceptable waiting times at bus interchange points are often contradictory problems because attempts to improve the coordination of bus services might deteriorate the individual performance metrics of each bus service. For instance, a bus service *A* can improve its EWT by re-scheduling its planned operations until the EWT of passengers is minimized. The same holds true for another bus service *B* which intersects with service *A* at a common bus stop *T* (transfer station). However, the optimal schedules that minimize the EWTs of services *A* and *B* might lead to significant waiting times at the transfer stop *T* for the passengers who are willing to transfer from bus service *A* to service *B* and vice versa. If one wants to reduce the transfer

waiting times at the stop T for improving the coordination between the two services, the services A and B should be re-scheduled and, consequently, they will no longer be optimal regarding the passengers' EWT. Therefore, it is evident that the coordination of various bus services is a trade-off between the reduction of waiting times at the transfer stops and the minimization of the passengers' EWT of each specific service.

2 Related studies

In the work of Ceder et al. (2001) bus timetables were re-scheduled for achieving maximal synchronization and the problem was formulated as a mixed integer programming one. However, the synchronization was achieved by simultaneous arrivals at transfer stations without considering the implications to the performance of each bus line. Daduna and Voß (1995) and Domschke (1989) worked also on scheduling timetables with a specific emphasis on synchronizing the transfers among different lines. The perplexities of such task included the headway differences among different lines that complicate the harmonization of passenger transfers. For this reason, Daduna and Voß (1995) adopted a multi-objective approach where the minimization of the sum of all transfer waiting times and the operating costs were considered by a heuristic optimization approach while an upper threshold for the maximum allowed waiting time was adopted.

In their work, Bookbinder and Desilets (1992) tried to reduce the transfer times of passengers by modifying the scheduled departure times of bus trips. For selecting the optimal departure time changes, a simulation-based approach was employed for considering the stochastic nature of the actual travel times of bus trips. Voss (1992) focused also on minimizing the transfer times of passengers in transit systems (buses and trains). The decision variables of the transfer time minimization problem were the departure time changes of scheduled trips (that affect the arrival times of trips at the transfer stops) and the problem was considered at its simpler version where all transit lines are met only at certain points and its more complicated version where they partially use the same tracks.

Jansen et al. (2002), Wong et al. (2008), Vansteenwegen and Van Oudheusden (2006) worked also on the problem of reducing the waiting time of passengers at the transfer stops while keeping the daily trips evenly spaced. Similarly to the approach of Bookbinder and Desilets (1992), the more recent studies of Cevallos and Zhao (2006a) and Cevallos and Zhao (2006b) tried also to reduce the complexity of the problem by merely shifting the departure times of the scheduled trips for finding the optimal solution for the passenger transfers with the use of a genetic algorithm.

Apart from the synchronization of the bus timetables, Zhao et al. (2003) focused on the more computationally complex problem of real-time coordination via a decentralized strategy where stops and buses communicate in real-time to achieve dynamic bus holding coordination at a local level.

There is a series of works also on bus and rail operations coordination. Coffey et al. (2012) optimized the time schedules of public transport modes for reducing missed connections under a multi-modal setting by trying to match the passenger demand expressed via journey planners with the public transport supply. Chien and

Schonfeld (1998), Sun and Hickman (2004), Shrivastava and O'Mahony (2009), Sivakumaran et al. (2012), Verma and Dhingra (2006) proposed multi-modal coordination approaches by considering the 'feeder model' where the bus system feeds a rail transit line and bus services have to adjust their schedules to the rail schedules. Nevertheless, the 'feeder line' concept is not applicable if the system is not multi-modal, but contains only bus transfers which are expected to have almost similar importance/weight.

This is also the direction of our work which models the bus coordination problem considering the passenger waiting times at transfer stations and the adherence to the KPIs of each bus service. The motivation behind our work is the non-evenly distributed passenger waiting times at transfer stations that can penalize the total passenger travel times even if the regularity of single bus services is optimal. In Sect. 3 the bus coordination problem is formulated considering the waiting times of passengers at transfer stations, the regularity-based KPIs of bus services (passenger EWT) and the operational constraints. The modeled bus coordination problem has exponential computational complexity and is computationally intractable even for small scenarios with two bus lines. For this reason, in Sect. 4 we present a heuristic solution space search strategy for converging to a nearly optimal solution that improves coordination by re-scheduling the departure times of the scheduled bus trips. Experimental results from two high-frequency, bi-directional services in Stockholm are presented in Sect. 5 showcasing a $\sim 13\%$ transfer waiting time improvement for only 2.8% deterioration of service regularity.

3 Modeling the bus coordination problem

Let us assume that $Q = \{Q_1, Q_2, \dots, Q_j, \dots, Q_{|Q|}\}$ are the bus services on a city network, where $|Q|$ is the total number of available bus services. Each bus service $Q_j \in Q$ operates a number of scheduled trips $R_{Q_j} = \{R_{1,Q_j}, R_{2,Q_j}, \dots, R_{j,Q_j}, \dots, R_{|R_{Q_j}|,Q_j}\}$ where $|R_{Q_j}|$ is the total number of daily trips for bus service Q_j . R_{Q_j} is already defined from the tactical planning phase of the broader Transit Network Planning problem (TNP) including the network design, timetable development and crew/vehicle scheduling for each bus line (Kepaptsoglou and Karlaftis 2009; Farahani et al. 2013; Ceder 2007; Gkiotsalitis and Cats 2017). The tactical planning phase defines also a number of operational constraints such as the expected service frequency during different daily periods (i.e., morning peak, afternoon peak) for adjusting to the passenger demand variations during the day. Let us assume that for each bus service $Q_j \in Q$ the solution of the TNP has defined a frequency range set:

$$F_{Q_j} = \begin{cases} F_{1,Q_j} = (a_1, b_1) \\ F_{2,Q_j} = (a_2, b_2) \\ \dots \\ F_{|F_{Q_j}|,Q_j} = (a_{|F_{Q_j}|}, b_{|F_{Q_j}|}), \end{cases}$$

where $F_{1,Q_j} = (a_1, b_1)$ is the minimum and maximum allowed departure time difference between consecutive bus trips of the same service during the first time period of the day for bus service F_{Q_j} and $|F_{Q_j}|$ is the number of time period splits within a day that enjoy different frequency settings (i.e., F_{1,Q_j} : morning peak, F_{2,Q_j} : midday etc.).

Each bus service Q_j serves also a number of bus stations $S_{Q_j} = \{S_{1,Q_j}, S_{2,Q_j}, \dots, S_{|S_{Q_j}|,Q_j}\}$ where $|S_{Q_j}|$ is the total number of bus stations served by service Q_j . In dynamic control operations, the bus service operator can re-schedule the bus service departure times (timetable re-scheduling). Given the bus travel times from one bus station to another during different day periods, one can calculate the *arrival time* of a bus trip at every bus station:

$$t_{R_{Q_j}, S_{|S_{Q_j}|}, Q_j} = t_{R_{Q_j}, S_{1,Q_j}} + \sum_{i=S_{1,Q_j}}^{S_{|S_{Q_j}|}, Q_j} \tau_{i, R_{Q_j}} \quad (1)$$

where $\tau_{i, R_{Q_j}}$ is the expected travel time of bus trip R_{Q_j} between two consecutive bus stations i to $i + 1$ plus the associated dwell time, $t_{R_{Q_j}, S_{|S_{Q_j}|}, Q_j}$ the arrival time of bus trip R_{Q_j} at the terminal and $t_{R_{Q_j}, S_{1,Q_j}}$ the departure time of bus trip R_{Q_j} from the origin that can be re-scheduled. Let $x_{R_{Q_j}}$ be the decision variable of the problem which represents the time variation of departure time $t_{R_{Q_j}, S_{1,Q_j}}$ after the re-scheduling and receives integer values (those integer values represent the time changes in minutes), then the arrival time of the bus trips at every station after re-scheduling is:

$$t_{R_{Q_j}, S_{|S_{Q_j}|}, Q_j}(x) = t_{R_{Q_j}, S_{1,Q_j}} + x_{R_{Q_j}} + \sum_{i=S_{1,Q_j}}^{S_{|S_{Q_j}|}, Q_j} \tau_{i, R_{Q_j}}. \quad (2)$$

For coordinating two bus services Q_i, Q_j , first there should be a search for the common interchange (transfer) stations. The transfer bus stations of two bus services are denoted as TS_{Q_i, Q_j} where $TS_{Q_i, Q_j} = S_{Q_i} \cap S_{Q_j}$:

$$TS_{Q_i, Q_j} = \{S_{1,Q_i}, S_{2,Q_i}, \dots, S_{|S_{Q_i}|}, Q_i\} \cap \{S_{1,Q_j}, S_{2,Q_j}, \dots, S_{|S_{Q_j}|}, Q_j\}. \quad (3)$$

Similarly, for coordinating three or more bus services, the transfer stations are: $TS_{Q_i, Q_j, Q_k} = (S_{Q_i} \cap S_{Q_j}) \cup (S_{Q_i} \cap S_{Q_k}) \cup (S_{Q_j} \cap S_{Q_k})$. Considering the operations of each service line, EWT is the performance index for regularity-based high frequency

bus services operating in dense metropolitan areas. EWT is calculated at each station $\{S_{1,Q_j}, S_{2,Q_j}, \dots, S_{|S_{Q_j}|,Q_j}\}$ of a bus service Q_j over the day for identifying the level

of service irregularity by calculating the excess waiting times (EWTs) of passengers. The EWT on a bus station S_{k,Q_j} for bus service Q_j is:

$$EWT_{S_{k,Q_j}}(x) = \frac{\left(\sum_{j=R_{1,Q_j}}^{R_{|R_{Q_j}|-1,Q_j}} (t_{j+1,S_{k,Q_j}} + x_{j+1} - t_{j,S_{k,Q_j}} - x_j)^2 \times 1/2 \right)}{(t_{R_{|R_{Q_j}|},S_{k,Q_j}} + x_{R_{|R_{Q_j}|}} - t_{R_{1,Q_j},S_{k,Q_j}} - x_{R_{1,Q_j}})} \quad (4)$$

$$- \frac{(t_{R_{|R_{Q_j}|},S_{k,Q_j}} + x_{R_{|R_{Q_j}|}} - t_{R_{1,Q_j},S_{k,Q_j}} - x_{R_{1,Q_j}})}{(2 \times (|R_{Q_j}| - 1))}$$

where $\sum_{j=R_{1,Q_j}}^{R_{|R_{Q_j}|-1,Q_j}} (t_{j+1,S_{k,Q_j}} + x_{j+1} - t_{j,S_{k,Q_j}} - x_j)$ is the summary of all headways between consecutive trips of service Q_j at station S_{k,Q_j} during the entire day. The first term of Eq. 4 represents the observed average waiting time of passengers at the bus station over a time period and the second the scheduled waiting time. In most cases, PTAs give different importance to the EWT of bus stations when computing the total EWT of the day for one bus service Q_j by adding weight factors at each bus station according to its importance. As a result, the EWT of one service Q_j is:

$$EWT_{Q_j}(x) = \sum_{i=S_{1,Q_j}}^{S_{|S_{Q_j}|,Q_j}} w_i \times EWT_i(x) \quad (5)$$

where w_i is the associated EWT weight to each station $i = \{S_{1,Q_j}, S_{2,Q_j}, \dots, S_{|S_{Q_j}|,Q_j}\}$.

If two or more bus services Q_i, Q_j intersect at one or more bus stations $TS = \{S_{1,Q_i}, S_{2,Q_i}, \dots, S_{|S_{Q_i}|,Q_i}\} \cap \{S_{1,Q_j}, S_{2,Q_j}, \dots, S_{|S_{Q_j}|,Q_j}\}$, then the coordination of services requires the optimization of waiting times at the transfer stations while also minimizing the EWT for each independent service Q_i, Q_j subject to tactical planning constraints such as the required bus frequencies. For passenger transfers from bus trips of service Q_j to bus trips of service Q_i , the waiting time at each transfer station for each trip $R_{x,Q_j} \in R_{Q_j}$ of service Q_j is:

$$WT_{R_{x,Q_j},TS_j}(x) = \min_{R_{y,Q_i} \in R_{Q_i}} \left\{ (t_{R_{y,Q_i},TS_j} + x_{R_{y,Q_i}}) - (t_{R_{Q_j},TS_j} + x_{R_{Q_j}}) \right\} \quad (6)$$

subject to : $(t_{R_{y,Q_i},TS_j} + x_{R_{y,Q_i}}) - (t_{R_{Q_j},TS_j} + x_{R_{Q_j}}) \geq 0.$

The trip R_{y,Q_i} that minimizes Eq. 6 is the trip from service Q_i that arrives at the transfer bus station TS_j after trip R_{x,Q_j} and its arrival time is closer to trip R_{x,Q_j} than any other trip $R_{y,Q_i} \in R_{Q_i}$. Computing the passenger waiting time of every trip R_{x,Q_j} at each transfer station TS_j requires to find first its closest trip from the connected bus service Q_i and this needs $|R_{Q_i}|$ computations. Equation 6 assumes that both services intersect at the same bus stop and does not consider service intersections where passengers have to walk from one bus stop to another as presented in Fig. 1. For computing the waiting time of all bus trips of service Q_j a number of $|R_{Q_j}||R_{Q_i}|$ computations is required and for expanding the computation to all bus transfer stations TS , the total number of computations is $|TS||R_{Q_j}||R_{Q_i}|$. The transfer waiting time at all transfer bus stations $TS_j \in (S_{Q_i} \cap S_{Q_j})$ during the day between the bus services Q_j, Q_i

is:

$$WT_{Q_j, Q_i}(x) = \sum_{k=TS_1}^{k=TS_{|TS|}} \sum_{\rho=R_{1,Q_j}}^{R_{|R_{Q_j}|,Q_j}} \min_{\forall R_{y,Q_i} \in R_{Q_i}} \left\{ t_{R_{y,Q_i},k} + x_{R_{y,Q_i}} - t_{\rho,k} - x_{\rho} \right\} \quad (7)$$

subject to: $t_{R_{y,Q_i},k} + x_{R_{y,Q_i}} - t_{\rho,k} - x_{\rho} \geq 0$

Concurrently, the waiting time of passenger transfers from service Q_i to service Q_j is:

$$WT_{Q_i, Q_j}(x) = \sum_{k=TS_1}^{k=TS_{|TS|}} \sum_{\rho=R_{1,Q_i}}^{R_{|R_{Q_i}|,Q_i}} \min_{\forall R_{y,Q_j} \in R_{Q_j}} \left\{ t_{R_{y,Q_j},k} + x_{R_{y,Q_j}} - t_{\rho,k} - x_{\rho} \right\} \quad (8)$$

subject to: $t_{R_{y,Q_j},k} + x_{R_{y,Q_j}} - t_{\rho,k} - x_{\rho} \geq 0$

From Eqs. 7, 8 one can derive the total transfer waiting time from all transfer stops $TS_1, \dots, TS_{|TS|}$ of two bus services Q_i, Q_j by adding the expected transfer waiting times at each transfer stop. However, a simple aggregation of the transfer waiting times at the transfer stops is not an adequate metric for measuring the total transfer waiting times of passengers since some transfer stops might accommodate significantly more transfers than others. For considering the demand for transfers, the

Fig. 1 Representation of two bus lines with common transfer stations

transfers at each stop $TS_e \in TS$ should have different importance. For realizing that, an importance weight $IW_{TS_e}^{Q_i, Q_j} = D_{TS_e}^{Q_i, Q_j} / (\sum_{k=TS_1}^{k=TS_{|TS|}} D_{TS_k}^{Q_i, Q_j})$ is assigned at every stop TS_e that accommodates transfers between services Q_i and Q_j where $D_{TS_e}^{Q_i, Q_j}$ is the number of daily transfers from service Q_i to Q_j at transfer stop TS_e . This way, the sum of the importance weights of all transfers between two services are equal to one: $\sum_{k=TS_1}^{k=TS_{|TS|}} IW_{TS_k}^{Q_i, Q_j} = 1$ and some transfer stops might have higher importance value (i.e., $IW_{TS_e}^{Q_i, Q_j} = 0.7$) while others a lower one (i.e., $IW_{TS_f}^{Q_i, Q_j} = 0.1$) indicating a lower number of transfers. The same approach of defining weight factors to bus stops for including the importance of the passenger demand to the waiting times is followed also by the transport authorities that use operational KPIs such as the EWT (refer to Eq. 5). Those importance weights are used to calculate the total transfer waiting times with the consideration of transfer demand after updating Eq. 7:

$$WT_{Q_i, Q_j}(x) = \sum_{k=TS_1}^{k=TS_{|TS|}} IW_{TS_i}^{Q_i, Q_j} \sum_{\rho=R_1, Q_i}^{R_{|R_{Q_i}|}, Q_i} \min_{\forall R_{y, Q_i} \in R_{Q_i}} \left\{ t_{R_{y, Q_j}, k} + x_{R_{y, Q_j}} - t_{\rho, k} - x_{\rho} \right\} \quad (9)$$

subject to: $t_{R_{y, Q_j}, k} + x_{R_{y, Q_j}} - t_{\rho, k} - x_{\rho} \geq 0$.

The coordination of two (or more) bus services Q_i, Q_j is a multi-objective problem with the following conflicting priorities: (1) minimize the EWT for each independent service Q_i, Q_j ; (2) minimize the passenger waiting times at the transfer stations. Therefore, a trade-off between the improvement of each service and the improvement of passenger transfer times should be established. For this reason, one can utilize weight factors that indicate the importance of each problem objective resulting in the formation of a single scalar objective function. The decision variables of the optimization problem are the rescheduling departure times x of each bus trip. For instance, when coordinating two bus services Q_i, Q_j with R_{Q_i} and R_{Q_j} trips our decision variables is a vector x of $|R_{Q_i}| + |R_{Q_j}|$ elements. If we consider also that from the frequency planning phase all consecutive bus trips of service Q_i must have a departure time difference in the range (a_1, b_1) minutes and of service Q_j in the range of (a_2, b_2) minutes for one time period of the day, then the coordination problem can be defined as:

$$\begin{aligned}
& \arg \min_x (x) = W_1 \times EWT_{Q_i}(x) + W_2 \times EWT_{Q_j}(x) + W_3 \times (WT_{Q_j, Q_i}(x) + WT_{Q_i, Q_j}(x)) \\
& \text{subject to } c_1(x) = t_{R_{2, Q_j}, S_{1, Q_j}}(x) - t_{R_{1, Q_j}, S_{1, Q_j}}(x) - a_1 \geq 0 \\
& \dots \\
& c_{|R_{Q_j}|}(x) = t_{R_{|R_{Q_j}|, Q_j}, S_{1, Q_j}}(x) - t_{R_{|R_{Q_j}|-1, Q_j}, S_{1, Q_j}}(x) - a_1 \geq 0 \\
& c_{1+|R_{Q_j}|}(x) = b_1 - (t_{R_{2, Q_j}, S_{1, Q_j}}(x) - t_{R_{1, Q_j}, S_{1, Q_j}}(x)) \geq 0 \\
& \dots \\
& c_{2|R_{Q_j}|}(x) = b_1 - (t_{R_{|R_{Q_j}|, Q_j}, S_{1, Q_j}}(x) - t_{R_{|R_{Q_j}|-1, Q_j}, S_{1, Q_j}}(x)) \geq 0 \\
& c_{2|R_{Q_j}|+1}(x) = t_{R_{2, Q_i}, S_{1, Q_i}}(x) - t_{R_{1, Q_i}, S_{1, Q_i}}(x) - a_2 \geq 0 \\
& \dots \\
& c_{2|R_{Q_j}|+|R_{Q_i}|}(x) = t_{R_{|R_{Q_i}|, Q_i}, S_{1, Q_i}}(x) - t_{R_{|R_{Q_i}|-1, Q_i}, S_{1, Q_i}}(x) - a_2 \geq 0 \\
& c_{2|R_{Q_j}|+|R_{Q_i}|+1}(x) = b_2 - (t_{R_{2, Q_i}, S_{1, Q_i}}(x) - t_{R_{1, Q_i}, S_{1, Q_i}}(x)) \geq 0 \\
& \dots \\
& c_{2|R_{Q_j}|+2|R_{Q_i}|}(x) = b_2 - (t_{R_{|R_{Q_i}|, Q_i}, S_{1, Q_i}}(x) - t_{R_{|R_{Q_i}|-1, Q_i}, S_{1, Q_i}}(x)) \geq 0 \\
& \forall x_i \in \mathbb{Z},
\end{aligned} \tag{10}$$

where W_1, W_2, W_3 are weight factors denoting the importance of minimizing the EWT times at services Q_i, Q_j and the importance of reducing the waiting times of transfer passengers. This problem is a combinatorial problem because the minute changes of the departure times for each trip, x_i , can take any integer value from a set q that contains all integer numbers since further denomination of timetable departure times is not acceptable from transport authorities for practical reasons. For instance, a typical timetable (general transit feed specification (GTFS) schedule) of a bus service is following the minutes' denomination. The denomination cannot be reduced to seconds since it is difficult to communicate this information to the bus driver who has to run the service. As might be expected, a further denomination to half-minutes can be applied (i.e., $x_i = 4.5$ min) during the operational control. In such case, the solution set will remain discrete and the options for departure time changes will be increased. The objective function of the problem is nonlinear and non-convex while all inequality constraints resulting from the frequency range limits are linear. Because of the non-convexity, it should be noted that the convergence to the global optimum cannot be guaranteed even if the problem was continuous by nature.

The above problem is a nonlinear integer programming problem because the decision variables can take discrete values. A solution method that is widely used for discrete optimization is the Branch and Bound (B&B) method (Land and Doig 1960) that relaxes the discrete optimization problem of Eq. 10 to a series of continuous sub-problems. Given the nonlinearity of the objective function, any continuous sub-problem of Eq. 10 can be solved with sequential

quadratic programming (SQP) which is one of the most effective numerical optimization methods for nonlinearly constrained optimization problems (Powell 1978). It cannot be guaranteed, though, that the solution of the SQP is a globally optimal solution of the problem at hand, notwithstanding the fact that SQP converges always to a locally optimal point after starting from an initial solution guess, because the objective function of Eq. 10 is non-convex; thus, it cannot be proved that one of the locally optimal SQP solutions is also a globally optimal one.

Due to that, it is required to use exact optimization algorithms that are more computationally expensive (such as the brute-force method) in order to guarantee the convergence to a globally optimal solution. The brute-force search evaluates the performance of all possible solutions. Finding a globally optimal solution with brute-force requires an exponential number of computations $|TS||R_{Q_i}||R_{Q_j}||q|^{|x|}$ where $|x|$ is the number of decision variables (in the case of Q_i, Q_j bus services, $|x| = |R_{Q_i}| + |R_{Q_j}|$).

Proof Let us consider the coordination of two bus services, Q_i, Q_j . If we modify the departure time of one trip the required computations for finding the consecutive bus trips that can be used for a potential transfer at any transfer stop are $|R_{Q_j}||R_{Q_i}|$ (according to Eq. 6). If the number of transfer stops between services Q_i and Q_j is $|TS|$, this cost grows to $|TS||R_{Q_j}||R_{Q_i}|$ for computing the transfer waiting times of passengers at all transfer stops after the departure time modification of the examined bus trip. Therefore, for each departure time modification a number of $|TS||R_{Q_j}||R_{Q_i}|$ computations is required for re-evaluating the waiting times of passengers at transfer stops. Each examined trip can modify its departure time by receiving any potential value from the set q . This results in a number of $|TS||R_{Q_j}||R_{Q_i}||q|$ computations for evaluating the impact of different departure times of an examined trip at the coordination problem of Eq. 10. Now, if one evaluates the full list of potential combinations of departure time modifications for all trips; then, the number of combinations that should be explored is $|q|^{|R_{Q_j}|+|R_{Q_i}|} = |q|^{|x|}$. Given the number of $|TS||R_{Q_j}||R_{Q_i}|$ computations which are required for evaluating the performance of each departure time modification, the number of required computations for evaluating all potential departure time modification options and selecting the globally optimal one is $|TS||R_{Q_j}||R_{Q_i}||q|^{|x|}$. $\square$

This means that a globally optimal solution cannot be computed even for very small scenarios by evaluating all possible departure time modification combinations. As an example, finding the global optimum re-scheduling solution with simple enumeration for two bus services with $|R_{Q_j}| + |R_{Q_i}| = 500$ daily trips requires $|TS||R_{Q_j}||R_{Q_i}||q|^{|x|} = |TS||R_{Q_j}||R_{Q_i}||q|^{500} \sim +\infty$ computations.

4 Coordinating bus services with sequential hill climbing after approximating all constraints with exterior point penalties

Due to the computational intractability of the bus coordination problem, we introduce a greedy method based on the hill climbing heuristic search for finding a solution for the bus coordination problem. It cannot be guaranteed though that the solution provided by the proposed heuristic is a globally optimal solution of the bus coordination problem. In addition, its convergence rate to a globally optimal solution cannot be established because a globally optimal solution cannot be calculated with exact numerical optimization methods for realistic problem instances. As a result, the performance of the solution provided by the hill climbing heuristic search is compared against the performance of other heuristic methods.

In the beginning, we introduce an *exterior point penalty function*, $p(x)$, to approximate the constrained bus coordination problem by an unconstrained one which is structured in a way that its minimization favors the satisfaction of the constraints. This penalty function adds to the objective function $f(x)$ a term that produces a high cost for the violation of any frequency range constraint; thus, providing a direction towards constraints' satisfaction. In this way, the constrained bus coordination problem of bus services Q_i, Q_j in Eq. 10 is approximated by the following unconstrained one:

$$\begin{aligned} \underset{x}{\text{minimize}} \quad & p(x) = f(x) + W_c \sum_{i=1}^{2|R_{Q_j}|+2|R_{Q_i}|} (\min[-c_i(x), 0])^2 \\ & \forall x_i \in q \end{aligned} \quad (11)$$

The penalty function $p(x)$ is structured in such a way that it is equal to the objective function value $f(x)$ if at some point we reach a solution x for which all frequency range constraints are satisfied: $\sum_{i=1}^{2|R_{Q_j}|+2|R_{Q_i}|} (\min[-c_i(x), 0])^2 = 0$. The term $W_c \sum_{i=1}^{2|R_{Q_j}|+2|R_{Q_i}|} (\min[-c_i(x), 0])^2$ is the added term to the objective function of the constrained optimization problem and dictates that if a constraint $c_i(x)$ has negative score, then $\min[-c_i(x), 0] = -c_i(x)$ since the constraint is violated for the current set of variables x . In this case, the objective function $f(x)$ is penalized by the term

$(-c_i(x))^2$ which returns a progressively higher penalty the further from zero $c_i(x)$ is. If a constraint $c_i(x) \geq 0$ for the current set of variables x , then this constraint is satisfied and is not penalizing the objective function since $\min[-c_i(x), 0] = 0$. In this way, we allow the progressive penalization of violating constraints $c_i(x) < 0$ (exterior point penalization) and we do not penalize any satisfied constraint $c_i(x) \geq 0$ even if it is an active one $c_i(x) = 0$ (no interior point penalization). We should note here that any constraint violation should penalize significantly the penalty function. For this reason, we included also a weight term W_c at the penalty function which ensures that if the penalty term $\sum_{i=1}^{2|R_{Q_i}|+2|R_{Q_j}|} (\min[-c_i(x), 0])^2 > 0$, the penalty function is significantly affected.

For minimizing the penalty function score, an iterative solution space search is performed based on hill-climbing with the aim of finding repeatedly a new solution x that provides a lower penalty function score than the previous one until no better solution can be found (convergence). Although there is no guarantee that this is the global optimum solution due to the heuristic nature of the search, the algorithmic search tries to return a near optimal solution. Initially, at iteration $k = 0$ we start with a random selection of decision variable values $x^{k=0}$. We can start for example from the “do-nothing” scenario where no re-scheduling is performed yet and $x_i^k = 0, \forall x_i^k \in x^{k=0}$. Starting from $x^{k=0}$ which is the *initial guess* solution with penalty function score $p(x^{k=0})$ we try to explore new solutions and replace the existing one in a greedy manner. If we coordinate bus services Q_i, Q_j we have a set of trips $R_{Q_i \cup Q_j} = \{R_{1,Q_i}, R_{2,Q_i}, \dots, R_{|R_{Q_i}|, Q_i}\} \cup \{R_{1,Q_j}, R_{2,Q_j}, \dots, R_{|R_{Q_j}|, Q_j}\}$ that can be re-scheduled. The search starts by selecting the first trip $\rho \in R_{Q_i \cup Q_j}$ and tries to update the

current solution $x^{k=0}$ with a new solution that reduces the value of the penalty function. In this greedy local search we test all possible departure time variations $x_{\rho}^{k=0}$ from a set $q = \{-5, \dots, 0, \dots, +5\}$ and select the one that returns the lowest penalty function score.

The set $q = \{-5, \dots, 0, \dots, +5\}$ min. does not include all integer values, as it should be according to the problem definition of Eq. 10, but a small sub-set of them. After the end of each iteration this departure time change is adopted and we search for new departure time changes in future iterations. Therefore, for a bus trip we might have a calculated departure time change of + 3 min. from the first iteration, + 2 min. from the second and + 4 min. from the third resulting in a total departure time change of + 9 min. if we had three iterations until convergence which would have been higher than the upper bound of set q ($= + 5$ min.). Therefore, the departure time changes for each trip can be outside those bounds (in fact, they can be any integer number). However, we use those bounds $q = \{-5, \dots, 0, \dots, +5\}$ min. of departure time changes at each iteration for the reduction of the severe computational costs.

If we had allowed an unbounded departure time change for each trip at each iteration of the search, then the set q would have contained all integer numbers and the number of computations for finding the optimal departure time change would have prohibited the implementation of the algorithm. Therefore, we allow only the search of a discrete number of limited options from the set $q = \{-5, \dots, 0, \dots, +5\}$ min. at each iteration and the algorithm is considerably faster without loss of generality because if a bus trip needs to postpone its departure time significantly, this will be calculated sequentially via postponing it little by little at each iteration until we converge to the final solution which cannot be further improved.

After testing all possible values from the set q for trip ρ , we continue sequentially by performing the same procedure for trip $\rho + 1$ and this continues until we reach the last daily trip of $R_{Q_i \cup Q_j}$. Therefore, we have an updated solution $x^{k=1}$ where $p(x^{k=1}) \leq p(x^{k=0})$. After that, we update the departure times of all trips considering the departure time changes: $t_{R_{l,Q_j}, S_{1,Q_j}} \leftarrow t_{R_{l,Q_j}, S_{1,Q_j}} + x_{R_{l,Q_j}}^{k=1}, \forall R_{l,Q_j} \in R_{Q_j}$ and

$t_{R_{l,Q_i}, S_{1,Q_i}} \leftarrow t_{R_{l,Q_i}, S_{1,Q_i}} + x_{R_{l,Q_i}}^{k=1}, \forall R_{l,Q_i} \in R_{Q_i}$. After the first iteration a number of con-

straints is not satisfied if the penalty function value $p(x^{k=1})$ is still greater than the objective function score $f(x^{k=1})$. In any case we start a new iteration $k = 2$ and we perform exactly the same greedy local search for replacing the departure time values of each trip with newer ones that reduce the penalty function score. Then, we update again the departure times of all trips $t_{R_{l,Q_j}, S_{1,Q_j}} \leftarrow t_{R_{l,Q_j}, S_{1,Q_j}} + x_{R_{l,Q_j}}^{k=2}, \forall R_{l,Q_j} \in R_{Q_j}$ and

$t_{R_{l,Q_i}, S_{1,Q_i}} \leftarrow t_{R_{l,Q_i}, S_{1,Q_i}} + x_{R_{l,Q_i}}^{k=2}, \forall R_{l,Q_i} \in R_{Q_i}$. Iterations continue in a similar manner and

after some iteration, $k = k'$, we would reach a point where the penalty function score is equal to the objective function score. This ensures that all constraints are satisfied. From this point onward, any improvement to the penalty function will be a direct improvement to the objective function because before iteration $k = k'$ there was no guarantee that solution updates that reduced the penalty function score reduced also the objective function score.

Finally, after a number of iterations, $k = k'' > k'$, it will not be possible to improve the penalty function value any further. At this point, the hill-climbing-based solution space search has converged to a locally optimal solution and the search terminates. It should be noted here that it cannot be guaranteed that the resulting solution is a globally optimal one.

```

1: function HILL-CLIMBING HEURISTIC FOR EXTERIOR POINT PENALTY FUNCTION MINIMIZ-
   ATION( $p(x), f(x)$ )
2:   Set iteration  $k = 0$  and initial random guess  $x^{k=0}$ ;
3:   Set  $q = \{-5, \dots, 0, \dots, +5\}$  minutes;
4:   while new iterations reduce the penalty function score:  $p(x^k) \neq p(x^{k-1})$  do
5:     Set iteration  $k \leftarrow k+1$ ;
6:     Set  $x_i^k = 0, \forall i \in R_{Q_i \cup Q_j}$ ;
7:     Store the penalty function value  $\mathbf{p} \leftarrow p(x^k)$ 
8:     for  $\rho$  in range  $R_{Q_i \cup Q_j}$  do
9:       for  $m$  in range  $\{1, \dots, |q|\}$  do
10:        Set  $z \leftarrow x_\rho^k$  and  $x_\rho^k \leftarrow m$ ;
11:        if  $p(x^k) < \mathbf{p}$  then
12:          Set  $\mathbf{p} \leftarrow p(x)$ ;
13:        else
14:          Set back  $x_\rho^k$  to its previous value:  $x_\rho^k \leftarrow z$ 
15:        end if
16:      end for
17:    end for
18:    Update the departure times of each bus trip:
19:     $t_{R_l, Q_j, S_{1, Q_j}} \leftarrow t_{R_l, Q_j, S_{1, Q_j}} + x_{R_l, Q_j}^k, \forall R_l, Q_j \in R_{Q_j}$ ;
20:     $t_{R_l, Q_i, S_{1, Q_i}} \leftarrow t_{R_l, Q_i, S_{1, Q_i}} + x_{R_l, Q_i}^k, \forall R_l, Q_i \in R_{Q_i}$ ;
21:  end while
22:  return solution  $x^{k=k''}$  where  $p(x^{k=k''}) = f(x^{k=k''})$  if all constraints are satisfied
23: end function

```

The computational complexity of this heuristic search can be expressed with the use of the big O notation which is used for classifying algorithms according to how their running time grows as the input size grows. This heuristic search is expected to converge into a solution in polynomial time after performing $|TS||R_{Q_i}||R_{Q_j}||q||x|k''$

computations which depend on the number of iterations until convergence k'' . This computational complexity allows us to coordinate two or more bus services by proposing new departure times for the scheduled trips.

Proof Let us consider the coordination of two bus services, Q_i, Q_j . As it is already demonstrated, a number of $|TS||R_{Q_j}||R_{Q_i}|$ computations is required for evaluating the performance of a departure time modification. In the proposed hill-climbing search, the search is performed in a sequential manner by starting from the first trip and evaluating all potential time modification options from the set q resulting in $|TS||R_{Q_j}||R_{Q_i}||q|$ computations. After the first trip, the same approach is performed for all other trips $\rho \in R_{Q_i \cup Q_j}$ in a sequential order requiring $|TS||R_{Q_j}||R_{Q_i}||q||x|$ computations for establishing the departure time modifications of trips that reduce further the cost of the penalty function. After the end of this search, the established departure time modifications are considered as the incumbent problem solution and the same iteration is performed k'' times until it is observed that the penalty function

score cannot be improved any further resulting in a total number of $|TS||R_{Q_j}||R_{Q_i}||q||x|k''$ computations until termination. $\square$

From the above proof, it is evident that the computational cost of the algorithm is polynomial and does not depend only on the size of the problem, $|R_{Q_j}| + |R_{Q_i}|$, and the total number of transfer stops $|TS|$, but also on the required number of iterations until the penalty function cannot be improved any further: k'' . The number of required iterations until convergence can vary significantly when solving different problem instances (i.e., different coordination problems). For this reason, a typical approach in numerical optimization is the use of an upper bound that determines the number of maximum allowed iterations until convergence for ensuring the termination of the optimization process even in the case that a solution cannot be found for a specific problem instance. If one uses a threshold value, K_{max} , for the maximum number of iterations k'' , then the computational complexity at the worst-case scenario can be described with the use of the big O notation as $O(|TS||R_{Q_j}||R_{Q_i}||q||x|K_{max})$.

5 Case study of regularity-based bus coordination

5.1 Stockholm test case

In this study, we utilized GTFS data from Sweden for deriving the planned schedule of public transport modes for the period 13 February 2016–17 June 2016. The data includes the following files: agency.txt (5 KB), stop_times.txt (242 MB), calendar.txt (17 KB), stops.txt (6.384 MB), calendar_dates.txt (890 KB), transfers.txt (1.02 MB), routes.txt (218 KB), trips.txt (11.628 MB).

For deriving the planned schedules, a library was developed in Python 2.7. The library processes .txt files and converts/stores them to an SQL database. This facilitates data queries and enables web-based visualization of the public transport operations with the use of OpenStreetMap (via OpenLayers, an open-source JavaScript library: <http://www.openlayers.org/api/OpenLayers.js>).

In particular, we focused on coordinating two bi-directional central bus services (Q_1 is bus line 1 and Q_4 is bus line 4) in Stockholm. Each direction of one bi-directional service has another EWT score and set of constraints, but the service-wide EWT can be computed as the average of each direction's EWT. Bus service Q_1 comprises of direction 1 (Essingetorget to Stockholm Frihamnen) and direction 2 (Stockholm Frihamnen to Essingetorget), bus service Q_4 comprises of direction 1 (Gullmarsplan to Radiohuset) and direction 2 (Radiohuset to Gullmarsplan).

Figure 2 depicts the examined bus lines 1 and 4. In addition, the EWT at each bus station calculated from the planned arrival times of the GTFS data according to Eq. 4 is presented in Fig. 5. The EWT of passengers at each bus station is calculated from 2 p.m.–7:30 p.m. because during the planning frequency stage, the daily operations of services Q_1 , Q_4 were split in four time periods and 2:00 p.m.–7:30 p.m. was

Fig. 2 Examined bus lines in Stockholm—bus line 1 (light blue color) and bus line 4 (purple color) (colour figure online)

one of them to which a frequency range constraint was assigned from the transport authority.

The allowed upper bounds for the headways between successive bus trips at the departure station are defined by the transport operator during the frequency planning phase and are presented in Table 1 for each direction of services Q_1, Q_4 . Those upper bounds are directly derived from the GTFS trip schedules. During the coordination of services, the departure times of bus trips can vary but they have to satisfy the upper bounds. In addition, to ensure that two successive bus trips do not start at the same time, we introduce a lower headway bound of 1 min at the departure stop.

There are also five (5) transfer bus stations of the two bi-directional services $TS_{Q_1, Q_4} = (S_{Q_1} \cap S_{Q_4})$: Västerbroplan (ID: 7446154), Mariebergsgatan (ID: 7446095), Fridhemsplan T-bana (ID: 7421661), Jungfrugatan (ID: 7445964), Värtavägen (ID: 7445967). The waiting times of passengers at the five bus transfer

Table 1 Headway upper bounds at the departure stop for the successive bus trips of services Q_1, Q_4 on both directions for the time period 2:00 p.m.–7:30 p.m.

Service	Headway upper bounds (min)
Q_1 : direction 1	12
Q_4 : direction 1	10
Q_1 : direction 2	12
Q_4 : direction 2	9

stations according to the GTFS data are presented in Fig. 4a. In more detail, in Fig. 4a are presented: (1) the aggregated transfer waiting times of passengers from service Q_1 to Q_4 for direction 1 trips which is equal to $WT_{Q_1, Q_4}^{dir.1} = 717$ min for all daily trips according to Eq. 7; (2) the aggregated transfer waiting times from service Q_4 to Q_1 for direction 1 trips which is equal to $WT_{Q_4, Q_1}^{dir.1} = 1808$ min for all daily trips; (3) the transfer waiting times from service Q_1 to Q_4 for direction-2-trips? which is equal to $WT_{Q_1, Q_4}^{dir.2} = 711$ min for all daily trips; (4) the transfer waiting times from service Q_4 to Q_1 for direction-2-trips? which is equal to $WT_{Q_4, Q_1}^{dir.2} = 1525$ min for all

daily trips. Those total transfer waiting times include all waiting times at each of the five transfer stops. The values of those transfer waiting times are so high because they are the sum of all waiting times at transfer stops during the entire day (not the average waiting time values). From the above, the total daily transfer waiting times from service Q_1 to Q_4 are $WT_{Q_1, Q_4} = WT_{Q_1, Q_4}^{dir.1} + WT_{Q_1, Q_4}^{dir.2} = 1428$ min and from service Q_4 to Q_1 are $WT_{Q_4, Q_1} = WT_{Q_4, Q_1}^{dir.1} + WT_{Q_4, Q_1}^{dir.2} = 3333$ min. The big difference to the total transfer waiting time from service Q_4 to Q_1 is due to the lower frequency of service Q_1 .

For optimizing the regularity-based bus service operations of the circular service we utilized the *exterior point penalty function*, $p(x)$, and the *hill-climbing* heuristic search strategy described in Sect. 4 for converging to a near optimal solution. The convergence starting point is the planned GTFS schedule ($x_i = 0, \forall x_i \in x$). The planned GTFS schedule EWT scores and transfer waiting times of passengers for services Q_1 and Q_4 are presented in Table 2. In Table 2 the service-wide EWT_{Q_1} of service Q_1 is the average of the EWT values from direction 1 and direction 2 presented in Fig. 5; therefore, $EWT_{Q_1} = 1/2(EWT_{Q_1}^{dir.1} + EWT_{Q_1}^{dir.2}) = 1/2(1.133 + 1.102) = 1.118$ min. Simi-

larly, the service-wide EWT_{Q_4} of service Q_4 is $EWT_{Q_4} = 1/2(EWT_{Q_4}^{dir.1} + EWT_{Q_4}^{dir.2}) =$

$1/2(1.053 + 0.84) = 0.947$ min. The EWT score is significantly lower than the transfer waiting times value because the transfer waiting times are the sum of all waiting times during the day while the EWT is the average value of excess waiting times at stations. Thus, a direct comparison between them requires the use of weight factors. Consequently, the bus service coordination objective function $f(x) = W_1 \times EWT_{Q_1}(x) + W_2 \times EWT_{Q_4}(x) + W_3 \times (WT_{Q_1, Q_4}(x) + WT_{Q_4, Q_1}(x))$ takes the value $f(x) = 1 \times 1.118 + 1 \times 0.947 + 0.000433732 \times (1428 + 3333) = 4.13$ considering weight factors $W_1 = 1, W_2 = 1, W_3 = 0.000433732$ for placing the same importance to the EWT scores of each service and to the total daily transfer waiting times. Finally, the exterior point penalty function $p(x) = f(x) + W_c \sum_{i=1}^{2|R_{Q_4}|+2|R_{Q_1}|} (\min[-c_i(x), 0])^2$ had initial score $p(x) = 30.13$ because 26 bus trips violated the frequency range limits of Table 1 and started their trip without any departure time difference from their preceding bus trip.

Fig. 3 Penalty function changes until convergence with hill-climbing. The penalty function value converged to the objective function value after 404 changes and convergence achieved after 542 changes

Fig. 4 Passenger waiting times at every transfer station for each direction of services Q_1, Q_4 in minutes (for the time period 2:10 p.m.–7:30 p.m.). The consecutive transfer bus trips at the transfer stations are presented at the horizontal axis, the transfer bus stations at the vertical axis and the transfer waiting time at each station is presented with a color scheme ranging from deep blue (synchronized transfers with 0 min transfer waiting time) to deep red (12 min transfer waiting time)

The convergence of the hill-climbing search is implemented on a 2556 MHz processor machine with 1024 MB RAM and is presented in Fig. 3 showcasing the coordination objective function, $f(x)$, and penalty function, $p(x)$, improvements after re-scheduling the departure times of bus trips. In Fig. 3 it is demonstrated how the hill-climbing search finds new bus trip departure times that reduce the value of the penalty function. In the horizontal axis is the number of iterative departure time changes, c , that kept decreasing the penalty function score. For each departure time change that decreased the penalty function score, one can note the changes to the EWT , the WT and the coordination objective function f . At the 404th penalty function decrease the penalty function score is for the first time equal to the coordination objective function $p(x^{c=404}) = f(x^{c=404}) = 2.463$. After this point all frequency range constraints are satisfied and any decrease to the penalty function results in an equivalent decrease of the coordination objective function. Finally, the penalty function value stops to decrease further after the 542nd change and we terminate the search assuming that we have reached the approximate global optimum point. At convergence, $p(x^{c=542}) = f(x^{c=542}) = 2.216$, $EWT(x^{c=542}) = 0.706$ min and $W_3 \times WT(x^{c=542}) = 1.509$ or $WT(x^{c=542}) = 3480$ min.

The regularity improvement after optimization for services Q_1, Q_4 in terms of EWT reduction of passengers at stations is presented in Fig. 5. In addition, passenger waiting times at each transfer station after optimization are presented in Fig. 4b showcasing the more evenly distributed waiting times of passengers at transfer stations compared to the previous ones from the planned schedule presented in Fig. 4a.

The original schedules of bus services Q_1, Q_4 were manually optimized from the bus operator based on ad-hoc rules; therefore, we were able to improve further the service regularity by 50–70% as shown in Fig. 5. Nevertheless, if the original schedules of bus services Q_1, Q_4 were *optimized based on regularity*, then any re-scheduling attempt of reducing the waiting times at transfer stations would have deteriorated the regularity level of those services. For this reason, a sensitivity analysis is important to understand how the transfer waiting time reduction affects the regularity of services and find a critical point where any gains from transfer waiting times improvements are less than the losses due to deterioration of service regularity. For this reason, we changed the value of the transfer waiting times weight factor, W_3 , and optimized again the bus coordination problem starting from $W_3 = 0$ which returns the optimal service regularity without considering the transfer waiting times.

Results are summarized in Fig. 6 where for $W_3 = 0$ the EWT of passengers for both services took its minimum possible value of $EWT = 0.591$ min. At this point, the total transfer waiting times of passengers over the entire day was $WT = 4204$ min. After re-scheduling the bus operations by putting more emphasis on coordination, we reach a point where for $W_3 = 0.0002$ we have $EWT = 0.609$ min. and $WT = 3641$. In other words, for 2.8% deterioration of the regularity of services Q_1, Q_4 we reduced the transfer waiting times by $\approx 13\%$. After this critical point, reducing further the transfer waiting times deteriorates significantly the EWT of individual services leading to an EWT of 0.706 min. for $W_3 = 0.0004337$. This 2.8% reduction of the regularity for services Q_1, Q_4 is well-justified because the transfer waiting times are improved by $\sim 13\%$ and, according to the literature Rietveld et al.

Fig. 5 EWT at every station and service-level EWT for every service before and after the heuristic search re-scheduling for the afternoon EWT phase (time period 2:10 p.m.–7:30 p.m.)

Table 2 Summary results showcasing the improvement of service-wide average EWTs and total daily transfer waiting times for services Q_1 , Q_4 after bus operations re-scheduling with the heuristic hill-climbing search

	Planned schedule (GTFS)	Coordination with hill-climbing optimization
EWT_{Q_1} score	1.118 min	0.299 min (73.26% improvement)
EWT_{Q_4} score	0.947 min	0.407 min (56.95% improvement)
WT_{Q_1} score	1428 min	1192 min. (16.53% improvement)
WT_{Q_4} score	3333 min	2288 min (31.35% improvement)
Penalty function score $p(x)$	30.13	2.216
Coordination obj. function score $f(x)$	4.13	2.216 (46.34% improvement)
Violated freq. constraints $c_i(x) < 0$	26	0
Computational time	N/A	27 min

Fig. 6 Optimization of the regularity-based bus coordination problem for different values of weight factors W_3 showcasing the trade-off between service regularity and passengers waiting times at transfer stations

(2001), Iseki and Taylor (2010), the transfer waiting times are perceived as even higher than they really are (up to 2–3 times) from the commuters.

5.2 Comparison against SoA solution methods

The scope of this analysis is to test the proposed exterior point sequential hill climbing heuristic against other heuristics or analytical optimization methods for determining the trade-off between solution accuracy (in terms of global optimum approximation) and computational costs. This is very important because the regularity-based bus service coordination problem is computationally intractable and high accuracy heuristics are required.

At first, exhaustive exact optimization methods -such as simple enumeration- are rejected due to the exponential computational costs. A solution method that can theoretically approximate the global optimum is the Branch and Bound (B&B) that relaxes the discrete optimization problem of Eq. 10 to a series of continuous ones that can be solved with sequential quadratic programming (SQP) given the non-linearity of the objective function. However, Eq. 10 is also non-convex and in practical implementations the SQP method converged to local optima when trying to solve each instance of the continuous regularity-based bus coordination problems. Thus, a multi-start strategy where the SQP is implemented multiple times by utilizing different initial search points for exploring more local optima is adopted. With this strategy, multiple local optima were computed and

the best-performing one was selected; thus, increasing the chances that the best-performing local optimum is closer to a globally optimal solution (Fig. 7).

In addition, two other discrete optimization heuristics are tested. The first heuristic is a problem-tailored simulated annealing search. Starting from the penalty function of Eq. 11 we generate a randomly selected initial guess of departure time variations for every bus trip denoted as $x^{\{0\}}$. This initial solution guess returns a penalty function score $p(x^{\{0\}})$ and we perform linear cooling by setting the initial temperature to $Temp = 1500$ according to the following algorithm.

```

1: function SIMULATED ANNEALING WITH LINEAR COOLING( $p(x), f(x)$ )
2:   Set iteration  $k = 0$  and initial random guess  $x^{k=0}$ ;
3:   Set initial temperature to  $Temp = 1500$ ;
4:   Set linear cooling function  $a(Temp, k) = Temp - k$ ;
5:   while  $k < 1500$  do
6:     Generate Random solution  $x$ ;
7:     if  $p(x) < p(x^{\{0\}})$  then
8:        $x^{\{0\}} \leftarrow x$ ;
9:     else
10:      Set  $Prob = e^{(p(x^{\{0\}}) - p(x)) / Temp}$ ;
11:      if  $U[0, 1] < Prob$  where  $U[0, 1]$  is a random number uniformly distributed on
12:          $[0, 1]$  then
13:            $x^{\{0\}} \leftarrow x$ ;
14:         end if
15:       end if
16:        $k \leftarrow k + 1$ ;
17:        $Temp = a(Temp, k)$ ;
18:     end while
19:   return optimal solution  $x^{\{0\}}$ 
20: end function

```

Last, we test a problem-specific sequential Genetic Algorithm where the GA chromosome is represented by the elements of a random solution x (i.e., each $x_i \in x$). Initially, we start with a set of population generators (parents) and we generate offsprings after a sequential crossover and mutation phase according to the following algorithm.

```

1: function SEQUENTIAL GA( $p(x), f(x)$ )
2:   Set iteration  $k = 0$ ;
3:   Set  $N\_max$  as the number of population generators (parents);
4:   For each parent,  $y$ , generate a random chromosome  $x^{\{y\}}$ ;
5:   while  $k < max_k$  do
6:     Select a random couple of parents  $y_i, y_j$ ;
7:     Generate an offspring  $O$  which is equal to parent  $y_i$  if  $p(y_i) < p(y_j)$  or, otherwise,
       equal to  $y_j$ ;
8:     Set best performing offspring,  $BO$ , initially equal to  $O$ 
9:     for  $\rho$  in range  $R_{Q_i \cup Q_j}$  do
10:      Sequential Crossover: Randomly select one of the  $\rho^{th}$  elements of chromosomes
         $y_i$  and  $y_j$ ;
11:      Replace the  $\rho^{th}$  element of  $O$  with that element
12:      if  $p(O) < p(BO)$  then
13:        Set  $BO \leftarrow O$ ;
14:        Set  $O \leftarrow BO$ ;
15:      end if
16:      Sequential Mutation: Randomly select a number  $n$  from the set  $q$ ;
17:      Replace the  $\rho^{th}$  element of  $O$  with that element;
18:      if  $p(O) < p(BO)$  then
19:        Set  $BO \leftarrow O$ ;
20:        Set  $O \leftarrow BO$ ;
21:      end if
22:    end for
23:    Replace the worst performing parent of the couple with  $BO$ ;
24:     $k \leftarrow k + 1$ ;
25:  end while
26:  return the parent which has the lowest penalty function score from the list of gen-
    erated parents
27: end function

```

Those two heuristics were selected instead of differential evolution or particle swarm optimization because of the discrete nature of the optimization problem. After optimizing the problem of Eq. 11 with the above-mentioned methods, the results are summarized in Fig. 7. The left sub-figure of Fig. 7 presents the

Fig. 7 Summary results comparing the penalty function optimization and the computational costs of heuristic search methods

improvement of the initial penalty function score (which was equal to 30.13) after optimization with simulated annealing with linear cooling, sequential GA, B&B with multi-start SQP and sequential hill-climbing demonstrating the significantly better convergence of the last two solution methods to the global optimum. The right sub-figure of Fig. 7 presents the computational times of the above-mentioned solution methods in minutes where the B&B with multi-start SQP had clearly the slowest convergence because of the implementation of the SQP optimizer many times from different starting points.

6 Discussion

This work addressed the problem of bus operations re-scheduling for the coordination of regularity-based services. First, the regularity-based bus coordination problem was modeled considering the regularity of bus services expressed in passenger EWT at stations, the coordination level expressed in passenger waiting time at transfer stations and operational constraints in the form of frequency ranges. The resulting problem was a constrained, integer nonlinear program with non-convex objective function that hinders the implementation of exact optimization methods for realistic problem instances. For this, a sequential hill-climbing heuristic method was introduced for converging to an approximate global optimum after using an exterior point penalty function to transform the constrained problem to an unconstrained one with the use of penalty terms.

Experimental results in two bi-directional bus services in Stockholm showcased a 50–70% room for improvement in terms of improving the regularity of existing services. In addition, the coordination at the transfer stations was able to improve 15–30% the transfer waiting times of passengers. Finally, we optimized the two bi-directional services based solely on regularity for identifying which is the optimal regularity of services if they are not coordinated and what is the trade-off between regularity deterioration and coordination improvement. Results summarized in Fig. 6 showcased that with 2.8% of regularity deterioration the transfer waiting times can be improved by 12.77% while after that point the coordination gains outweigh the regularity losses.

In addition, the proposed sequential heuristic search for coordinating regularity-based bus services was tested against other heuristic search methods such as (1) simulated annealing with linear cooling, (2) sequential GA and (3) B&B with SQP and a multi-start strategy. The results of optimizing the operations of the same buses were summarized in Fig. 7. In particular, the simulated annealing with linear cooling search method had the worst performance because the generation mechanism of better performing elements was random and was not capable of exploring the solution space in an efficient way. The sequential GA search inherited was not a simple GA implementation, but inherited the advantageous characteristic of the sequential hill-climbing since the crossover and the mutation were conducted in a sequential approach via an element-by-element investigation. The computational cost of this method was higher compared to the sequential hill climbing due to the presence of more than one population generators and it did not return a good approximation to

the global optimum because it did not search all potential departure time change options at the sequential search but mainly relied on the mutation to explore more options. Finally, the best performing option of B&B with SQP and a multi-start strategy was significantly slower than all other methods. The reason was the implementation of the SQP optimizer many times from different starting points to compute more local optima in order to find the closest one to the global optimum due to the non-convexity of the objective function and the need to run the entire process repeatedly for computing different instances of the B&B method since the regularity-based bus coordination problem is a discrete one. Therefore, its practical implementation is hindered.

7 Concluding remarks

This work examined the problem of bus operations re-scheduling for the coordination of regularity-based bus services. Summarizing the results from the previous sections, in the concluding remarks we present the lessons learned from this study.

First, the excess waiting times of passengers of existing services can be improved by more than 50% by modifying the departure times of the scheduled trips without considering the synchronization of services. This is in line with the on-site results from the work of Strathman et al. (1999) that used limited dispatching time modifications to improve the service regularity in Portland (transit provider: Tri-met) demonstrating a practical benefit of up to 23% on the coefficient of variation of the bus headways even in the case where the recorded trip travel times from the daily operations differ significantly from the expected ones.

Second, the waiting times at the transfer stops were able to be improved by 15–30% at the examined two-line network of Stockholm demonstrating that modifying the scheduled timetables of trips can have a positive impact to the transfer waiting times.

Third, an optimal synchronization among lines that minimizes the transfer waiting times of passengers has a significantly negative impact to the service regularity because the service regularity should be reduced for achieving an improved synchronization. For instance, from the experiments we learned that the total transfer waiting times can be improved by up to $\sim 13\%$ by deteriorating the service regularity by $\sim 3\%$. Nevertheless, any further transfer waiting time improvement impacts the service regularity in a disproportionate manner.

Finally, the proposed sequential hill-climbing method has almost similar performance with the method of B&B that employs multi-start SQP and requires only a limited number of computations for converging to a nearly optimal solution; something that enables its use for larger-scale applications in contrast to the B&B which is impractical.

In future research, the proposed coordination method for regularity-based services can be expanded also to other modes (i.e., trains) which operate under a regularity scheme allowing the cooperation between multiple modes. Finally, one can argue that the capacity of the vehicles might have an impact on synchronized transfers if a synchronized transfer leads to overcrowding. However, this analysis requires

real-time, high-granularity data regarding the occupancy levels of buses and such case study can be explored in future research with a specific focus on real-time synchronization. In such case, the capacity of buses can be included as an inequality constraint at the modeling stage of the bus service coordination problem.

Open Access This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

References

- Bookbinder JH, Desilets A (1992) Transfer optimization in a transit network. *Transp Sci* 26(2):106–118
- Botea A, Braghin S (2015) Contingent versus deterministic plans in multi-modal journey planning. In: Twenty-fifth international conference on automated planning and scheduling
- Botea A, Nikolova E, Berlingiero M (2013) Multi-modal journey planning in the presence of uncertainty. In: Twenty-third international conference on automated planning and scheduling
- Cats O, Larijani A, Ólafsdóttir Á, Burghout W, Andreasson I, Koutsopoulos H (2012) Bus-holding control strategies: simulation-based evaluation and guidelines for implementation. *Transp Res Rec* 2274:100–108
- Ceder A (2007) Public transit planning and operation: theory. Modeling and practice. Butterworth-Heinemann, Oxford
- Ceder A, Golany B, Tal O (2001) Creating bus timetables with maximal synchronization. *Transp Res Part A Policy Pract* 35(10):913–928
- Cevallos F, Zhao F (2006a) A genetic algorithm for bus schedule synchronization. In: Applications of advanced technology in transportation, pp 737–742
- Cevallos F, Zhao F (2006b) Minimizing transfer times in public transit network with genetic algorithm. *Transp Res Rec* 1971:74–79
- Chien S, Schonfeld P (1998) Joint optimization of a rail transit line and its feeder bus system. *J Adv Transp* 32(3):253–284
- Chowdhury M, Chien S (2011) Joint optimization of bus size, headway, and slack time for efficient timed transfer. *Transp Res Rec* 2218:48–58
- Coffey C, Nair R, Pinelli F, Pozdnoukhov A, Calabrese F (2012) Missed connections: quantifying and optimizing multi-modal interconnectivity in cities. In: Proceedings of the 5th ACM SIGSPATIAL international workshop on computational transportation science. ACM, pp 26–32
- Daduna JR, Voß S (1995) Practical experiences in schedule synchronization. In: Computer-aided transit scheduling. Lecture Notes in Economics and Mathematical Systems 430, Springer, Berlin, pp 39–55
- Dessouky M, Hall R, Zhang L, Singh A (2003) Real-time control of buses for schedule coordination at a terminal. *Transp Res Part A Policy Pract* 37(2):145–164
- Domschke W (1989) Schedule synchronization for public transit networks. *Oper Res Spectr* 11(1):17–24
- Farahani RZ, Miandoabchi E, Szeto WY, Rashidi H (2013) A review of urban transportation network design problems. *Eur J Oper Res* 229(2):281–302
- Fu L, Yang X (2002) Design and implementation of bus-holding control strategies with real-time information. *Transp Res Rec* 1791:6–12
- Gkiotsalitis K, Cats O (2017) Exact optimization of bus frequency settings considering demand and trip time variations. In: 96th Transportation research board annual meeting
- Gkiotsalitis K, Maslekar N (2015) Improving bus service reliability with stochastic optimization. In: 2015 IEEE 18th international conference on intelligent transportation systems (ITSC). IEEE, pp 2794–2799
- Iseki H, Taylor BD (2010) Style versus service? An analysis of user perceptions of transit stops and stations. *J Public Transp* 13(3):2

- Jansen LN, Pedersen MB, Nielsen OA (2002) Minimizing passenger transfer times in public transport timetables. In: 7th Conference of the Hong Kong Society for Transportation Studies, Transportation in the information age, pp 229–239
- Kepaptsoglou K, Karlaftis M (2009) Transit route network design problem: review. *J Transp Eng* 135(8):491–505
- Land AH, Doig AG (1960) An automatic method of solving discrete programming problems. *Econometrica* 28:497–520
- Powell MJ (1978) The convergence of variable metric methods for nonlinearly constrained optimization calculations. *Nonlinear Program* 3:27–63
- Rietveld P, Bruinsma FR, Van Vuuren DJ (2001) Coping with unreliability in public transport chains: a case study for Netherlands. *Transp Res Part A* 35(6):539–559
- Shrivastava P, O'Mahony M (2009) Modeling an integrated public transportation system—a case study in Dublin, Ireland. *Eur Transp* 41:28–46
- Sivakumaran K, Li Y, Cassidy MJ, Madanat S (2012) Cost-saving properties of schedule coordination in a simple trunk-and-feeder transit system. *Transp Res Part A* 46(1):131–139
- Strathman JG, Dueker K, Kimpel TJ, Gerhart R, Turner P, Callas S, Griffin D (1999) Service reliability impacts of computer-aided dispatching and automatic vehicle location technology: a trimet case study. Published by the Center for Urban Studies, College of Urban and Public Affairs, Portland State University. Catalog Number: PR125
- Sun A, Hickman M (2004) Scheduling considerations for a branching transit route. *J Adv Transp* 38(3):243–290
- Vansteenwegen P, Van Oudheusden D (2006) Developing railway timetables which guarantee a better service. *Eur J Oper Res* 173(1):337–350
- Verma A, Dhingra S (2006) Developing integrated schedules for urban rail and feeder bus operation. *J Urban Plan Dev* 132(3):138–146
- Voss S (1992) Network design formulations in schedule synchronization. In: *Computer-aided transit scheduling. Lecture Notes in Economics and Mathematical Systems* 386, Springer, Berlin, pp 137–152
- Wong RC, Yuen TW, Fung KW, Leung JM (2008) Optimizing timetable synchronization for rail mass transit. *Transp Sci* 42(1):57–69
- Xuan Y, Argote J, Daganzo CF (2011) Dynamic bus holding strategies for schedule reliability: optimal linear control and performance analysis. *Transp Res Part B* 45(10):1831–1845
- Yu B, Yang Z, Sun X, Yao B, Zeng Q, Jeppesen E (2011) Parallel genetic algorithm in bus route headway optimization. *Appl Soft Comput* 11(8):5081–5091
- Zhao J, Bukkapatnam S, Dessouky MM (2003) Distributed architecture for real-time coordination of bus holding in transit networks. *IEEE Trans Intell Transp Syst* 4(1):43–51