Distributed train timetable synchronization in metro network: An ADMM-based decomposition framework

Renming Liu[a,c], Shukai Li[a], Lixing Yang[a], Ronghui Liu[b]

[a]*School of Systems Science, Beijing Jiaotong University, Beijing, 100044, China*
[b]*Institute for Transport Studies, University of Leeds, Leeds LS2 9JT, U.K.*
[c]*SF Express (Group) Co., Ltd., Shenzhen 518000, Guangdong, China*

Abstract

The increasing spatial or temporal scales of metro networks generate an important research challenge in developing fast and efficient optimization methods for handling the train timetable synchronization problem (TTSP). This paper develops a distributed optimization algorithm for the TTSP of complex metro networks, with the objective of minimizing both the waiting time of inbound and transferring passengers in the whole network. We construct explicit dynamic equations of train passenger loads throughout the network and quantify the transferring passengers at transfer stations. These equations encapsulate the dynamic passenger transfer behavior within the metro system. To deal with the computationally expensive large-scale MINP problem, an alternating direction method of multipliers (ADMM) based decomposition approach is proposed to split the original TTSP into a set of single-line timetabling subproblems that can be solved in a decentralized manner. Furthermore, a novel heuristic two-level ADMM-based approach, where the upper level decides the connections among trains of different lines and the lower level applies standard ADMM with fixed binary variables to optimize the timetable, is designed to deal with the nonconvexity issue. We demonstrate its ability to conveniently obtain a high-quality solution to the network timetable synchronization problem numerically.

Keywords: Urban metro network, Timetable sychronization, Distributed optimization, Alternating directions method of multipliers

1. Introduction

1.1. Background and motivations

In large metro networks, reducing transfer times through carefully coordinating train arrival times at stations would greatly improve the service quality as well as potentially enhance the operating capacity of the metro system. The train timetable synchronization problem (TTSP) has been receiving significant attention in the literature (Corman et al., 2014; Gkiotsalitis et al., 2023). TTSP aims to enhance the coordination between train services from different lines by controlling train arrival and departure time, in particular, by adjusting the headway time between trains, running time between stations, and dwell time at each station. Together with factors such as the passenger loads along the lines, and passenger routing variable, the problem becomes more complex (Cacchiani et al., 2014; Wu et al., 2019; Yuan et al., 2022). The ultimate goal of

*Corresponding author.
Email address: shkli@bjtu.edu.cn (Shukai Li)

timetable coordination is to synchronize the train arrival times at interchange stations so that passengers can transfer from line to line with minimum transfer time while no excess waiting time incurred at non-transfer stations. To this end, this paper proposes a timetable synchronization method using mixed integer programming optimization, in which the dynamic passenger loads, departure and arrival time evolution, and transfer waiting time constraints are explicitly considered. Solving such a centralized optimization for the whole network with multiple lines and trains imposes a high computational burden, and scales poorly with the number of lines and trains. Conversely, distributed optimization techniques allow for decomposing the original large-scale TTSP into a series of small-scale single-line subproblems, which can then be solved in a parallel manner. Furthermore, practical applications can also benefit from the scalability and robustness of the distributed optimization algorithm, as it is highly modularized and flexible. For these reasons, this paper proposes an advanced distributed optimization approach to address the large-scale TTSP.

1.2. Literature review

Timetabling for a metro line entails determining the train arrival and departure times at each station to satisfy a given service frequency that meets the travel demands (Assis and Milani, 2004; Corman et al., 2017; Zhang et al., 2019; Jin et al., 2022). Timetabling for a metro network would need to coordinate the arrivals and departures of trains at transfer stations to reduce passenger transfer delays. A large body of literature addresses this problem by adopting a variety of objective functions, including minimization of transfer waiting time, maximization of synchronization, and a combination of these different objectives. Schröder and Solchenbach (2006) applied the quadratic semi-assignment model to improve the amount of available transfer time together with the users' satisfaction. Measured to the original timetables, the adjustment leads to improvements between 0.5% and 5%. Wong et al. (2008) modeled the TTSP into a mixed integer programming (MIP) where the headway time is a decision variable and designed an effective heuristic method to obtain high-quality solutions. Another MIP model minimizing transfer delays by allowing extra dwell time was proposed by (Shafahi and Khani, 2010). Wu et al. (2015) considered a similar MIP model with the objective of minimizing the maximal transfer waiting time, aiming for preventing excess transfer waiting time. In parallel and for better operation efficiency, numerous research has been carried out to collaboratively optimize the train timetable with other aspects, such as delay management (Corman et al., 2017), train speed control (Ye and Liu, 2016, 2017) and vehicle scheduling (Fonseca et al., 2018).

Besides, many researchers have considered the TTSP with multi-objective functions. For example, Kwan and Chang (2008) maximized the total passenger satisfaction without deviating from the given timetable too much via setting headway time, layover time, dwell time, and running time as variables. The results, solved by a specially designed nondominated sorting genetic algorithm 2 (NSGA 2), show that the average transfer time of all interchange stations is reduced. Moreover, the synchronization is improved with more "just meet" and fewer "just miss". Ceder et al. (2013) tried to keep the headways as even as possible while minimizing the deviation from desired train passenger loads. By designing two tailored heuristics, the load discrepancy is successfully reduced from 38% to a deviation of less than 15% with an even-headway timetable. Niu et al. (2015) created a nonlinear timetabling model for balancing the passenger waiting time on platforms and crowding level onboard under time-variant passenger demands. The numerical experiments suggest the superiority of the genetic algorithm over the dynamic programming approach in the two-line case, resulting in reduced average passenger waiting time and crowding level. Wang et al. (2015) established an event-driven model with three objectives: energy consumption, total travel time, and last train waiting time. The used sequential quadratic programming method obtains a non-fixed headway timetable, which provides better performances than the fixed-headway timetable while consuming more computational time. Fonseca et al. (2018) proposed a MIP model combining the train timetabling and rolling stock scheduling problems together to optimize transfer costs and operational costs simultaneously. Numerical experiments using a special heuristic demonstrated a reduction in the average excess transfer time of 2.8-4.5 min and an increase in the number of ideal transfers by 175% compared to the practical timetable.

Moreover, there is a number of research contributing to the first or last train connection problem in metro networks. Kang et al. (2015b) focused on the last train rescheduling problem, aiming for maximum available transfer time and accessibility with minimum running and dwell time. In addition, Kang et al. (2015a) first considered the last train timetabling problem as an NLP, and further improved it as MILP in (Kang and Meng, 2017). Guo et al. (2016) studied a MILP model and a sub-networks connection method to improve the transfer efficiency for the first train timetabling problem, showing reduced network transfer waiting time in the numerical example.

However, most of the research attempted to either maximize the train connectivity or minimize the waiting time for transferring between train services with the assumption of constant transferring passengers (Wong et al., 2008; Wu et al., 2015; Guo et al., 2020). In reality, transfer passengers vary dynamically with headway adjustments and demand fluctuations, creating interdependencies between timetable decisions and passenger flowsa challenge rarely addressed in prior work. For instance, Blanco et al. (2020) proposed an integrated MILP model for line planning and timetabling with time-dependent demand and explicitly modeled interchange passengers, their approach relies on piecewise linear approximations of passenger flows. Nguyen et al. (2021) modeled the dynamic transferring passengers relying on if-condition rules to capture transfer connectivity (e.g., "if a feeder train arrives within a time window, passengers can transfer"), bypassing explicit analytical coupling between passenger loads and timetable variables. Yin et al. (2023) discretized the time horizon and linearized nonlinear flow dynamics with auxiliary variables, focusing on time-period-based aggregation of passenger demand. In contrast, our model directly embeds continuous-time passenger dynamics into the optimization, rigorously enforcing transfer connectivity through binary variables to compute exact waiting times.

Besides, the existing works about train scheduling/rescheduling problems in large-scale metro networks are usually conducted in a centralized fashion (Corman et al., 2012; Kang et al., 2016). To reduce computation complexity, distributed optimization methods are developed to split the large-scale problem into sets of low-dimensional subproblems, which can be solved by a local processor (Boyd et al., 2011). The distributed optimization methods have been widely adopted for different transportation problems, including railway network design problem (Bärmann et al., 2017), vehicle speed or trajectory optimization in road networks (Tajalli and Hajbabaie, 2018; Mirheli et al., 2019), signal timing optimization for urban networks (Al Islam and Hajbabaie, 2017; Le et al., 2017), vehicle routing problem in corridor networks (Zhou et al., 2018), real-time vehicle dispatching problem in bus transit network (Zhao et al., 2003), and train speed control problems (Li et al., 2020). In contrast, the distributed optimization methods for large-scale train timetabling problems in metro networks have not been fully studied in the literature.

Motivated by these discussions, this paper proposes a distributed optimization algorithm, based on the alternating direction method of multipliers (ADMM), to adjust the timetable of individual lines for better train synchronization at transfer stations, while also controlling the waiting time of non-transferring passengers. ADMM takes advantage of both the decomposability of dual ascent and the convergent method of multipliers (Boyd et al., 2011; Wang and Ong, 2017). By constructing the augmented Lagrangian formulation, two primal variables are updated in an alternating manner (termed as alternating direction), where the dual variable is updated using the augmented Lagrangian parameter. In this study, the original optimization problem is non-convex due to the binary variables. Then the convergence can not be guaranteed if applying the standard ADMM directly. To provide a feasible suboptimal solution, we design a novel two-level approach that alternatively fixes the binary variables and the other variables and solves these subproblems in an iterative way. This two-level approach is able to improve the solution quality and provide a sub-optimal solution. Specifically, we list the main contributions of this study as follows:

(1) This paper constructs a network-wide train timetable synchronization model, which explicitly couples continuous-time passenger load dynamics with timetable variables through state equations, capturing exact transfer connectivity via binary variables, and avoiding heuristic rules (Nguyen et al., 2021) or demand periodization (Yin et al., 2023). The objective function is therefore more comprehensive since it takes both the waiting time for inbound and transferring passengers into consideration, promoting extra gains in the operation efficiency of metro systems. In the formulation, the systematic constraints on headway time and dwell time, passenger load dynamic, and transfer waiting time are established. The formulated model turns out to be a MINP model.

(2) The proposed complex MINP model of the TTSP has a relatively high computational burden even in a relatively small network, making it intractable to be directly solved by existing solvers for a large-scale network. By introducing global common variables and a consensus constraint, the original problem is rewritten as a global consensus problem (Boyd et al., 2011), which enables decentralized computation using standard ADMM. Due to the fundamental nonconvexity issue of the solution space, a two-level ADMM-based method is further proposed to acquire suboptimal solutions. In the upper level, a timetabling problem is handled by the standard ADMM with fixed binary variables (or linearly relaxed binary variables in the initialization step); while in the lower level, the binary variables that determine the train connections are solved by taking the fixed timetable from the upper level as input. This iterative and heuristic two-level method is numerically validated to be efficient. In contrast to centralized optimization approaches, this decomposition method can be conducted in parallel so as to achieve the effective reduction of computing time.

The rest of this paper is organized as follows. Section 2 presents the characteristics of train traffic in metro networks and formulates the timetable synchronization model. In Section 3, we develop a two-level ADMM-based approach to solve it. In Section 4, we validate the proposed model and method through several numerical experiments. Finally, Section 5 gives the summary of this paper.

2. Problem formulation

In this study, a TTSP in metro networks is investigated by considering the passenger load dynamic and transfer waiting time. To depict this problem, a train traffic model incorporates the safety and operational constraints to reflect the dynamics in headway time and passenger loads. Then the constraints for passenger transfer waiting time are constructed using a series of binary variables. Subsequently, in order to conveniently formulate a metro timetable synchronization optimization model and develop effective algorithms, some assumptions are given in the following:

Assumption 1. The transfer time $e_m^{m'}(k)$ which represents the walking time from Line m' to Line m at station k, is a known constant for all passengers.

Assumption 2. For simplicity, overtaking and crossing operations are not considered and all trains depart from the initial stations following the first-in-first-out rule. Each station can only accommodate one train at any time.

Assumption 3. In order to streamline the computation of transfer waiting times, we make the assumption that passengers always board the earliest arriving train.

Assumption 4. The number of alighting passengers is calculated as a proportion of onboard passengers, and the transferring passengers are a proportion of the alighting passengers. The proportion coefficients can vary with different train i at different station m, as well as time periods.

In **Assumption 1**, the transfer time $e_m^{m'}(k)$ stands for the walking time used by a passenger moving from the track of the feeder line m' to the track of the connecting line m, and their values can be estimated according to real-world statistical data (Nguyen et al., 2021). This assumption aligns with scenarios where passenger inflows are controlled or relatively low (e.g., off-peak hours (Guo et al., 2017; Yin et al., 2017) or last-train operations (Guo et al., 2020)), minimizing variability in walking speeds due to regulated platform densities. While station-specific walking times could enhance realism (e.g., layout-dependent speeds in Guo et al. (2021); Zhou et al. (2023); simulation-based calibration in Schmaranzer et al. (2020)), such approaches introduce complexity and often rely on heuristic or meta-heuristic methods (e.g., genetic algorithms, ALNS). **Assumption 2** is widely used in literature (Wang et al., 2015, 2017), and holds for most of the urban metro networks. **Assumption 3** requires that the train loading capacity always satisfies the passenger demands at each station, which may not be the case all the time. Nevertheless, many metro systems adopt inflow restrictions (e.g., fare gate controls, queue management) to prevent platform overcrowding (Li et al., 2017; Yuan et al., 2022). These measures ensure that passenger demand does not exceed train capacity, aligning with the assumption. Our model treats the passenger arrival rate as the regulated inflow rate entering platforms, thus this assumption holds valid, especially during off-peak hours when passenger demand is naturally lower. The same assumption is also adopted in Guo et al. (2017); Yin et al. (2017). Note that we can relax **Assumption 3** and extend this work to peak-hour scenarios using explicit congestion modeling (e.g., onboard crowding penalties or passenger inflow control). However, this extension requires additional constraints and objectives (e.g., minimizing passenger queueing times outside stations), which we defer to future investigation. Our current focus on off-peak hours allows us to rigorously address the core challenge of network-wide timetable synchronization without overcomplicating the model. For **Assumption 4**, one can estimate the proportionality factors using the statistical data of passenger demand (Li et al., 2017).

In large metro networks, TTSP aims to coordinate train arrivals at transfer stations to minimize passenger waiting times while respecting operational constraints (e.g., headway limits, dwell times). The problem becomes increasingly complex due to three factors: a) transferring passengers dynamically depend on headway adjustments, creating interdependencies between lines; b) binary variables representing train connections introduce nonconvexity complexity; c) centralized optimization methods do not scale efficiently with networks.

To address these challenges, we propose a distributed optimization framework that decomposes the network-wide problem into single-line subproblems. This approach explicitly models passenger load dynamics and leverages ADMM to iteratively coordinate solutions across lines while ensuring consensus at transfer stations.

2.1. Mathematical Formulations

2.1.1. The train traffic dynamics with single line constraints

For each line $m \in \mathcal{M}$, the evolution of train departure time from station k and arrival time at station $k+1$ should satisfy

$$TA_i^m(k+1) = TD_i^m(k) + r_i^m(k), \tag{1}$$

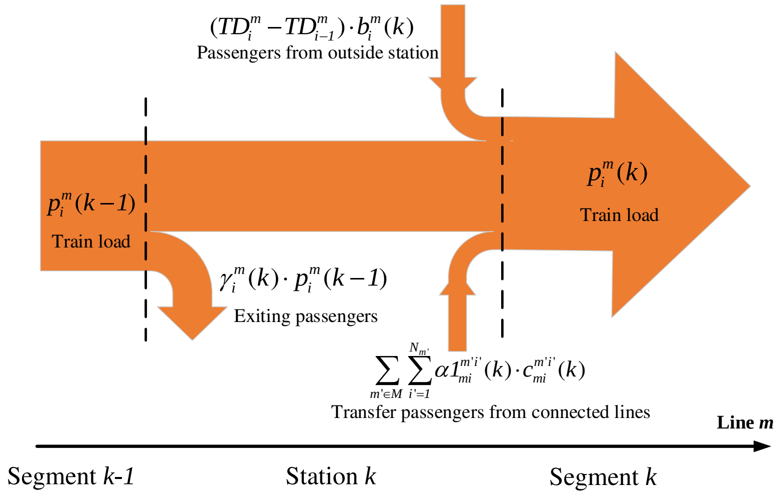

Figure 1: Illustration of passenger load dynamic of train i at station k.

where $TA_i^m(k)$ denotes the arrival time of train i at station k on Line m, $TD_i^m(k)$ is the departure time of train i from station k on Line m, and $r_i^m(k)$ is the running time of train i from station $k-1$ to k on Line m.

Due to the requirements of operation safety and service level, the headway (departure time intervals of successive trains) is bounded at each station by $H_{\min}^m$ and $H_{\max}^m$. Similarly, the dwell time is bounded by $S_{\min}^m$ and $S_{\max}^m$. Thus, for each $m \in \mathcal{M}$, $k = 1, 2, \ldots, K_m$, and $i = 1, 2, \ldots, N_m$,

$$\begin{aligned} H_{\min}^m &\leq TD_{i+1}^m(k) - TD_i^m(k) \leq H_{\max}^m, \\ S_{\min}^m &\leq TD_i^m(k) - TA_i^m(k) \leq S_{\max}^m. \end{aligned} \tag{2}$$

Equations (1) and (2) demonstrate that the headway time and dwell time change along with stations and trains, and $TA_i^m(k)$ and $TD_i^m(k)$ are decision variables. Particularly, for $i = 1$, the headway is set as the difference of station k's opening time, u_m, and train 1's departure time, i.e., $TD_1^m(1) - u_m$. For simplicity, the headway of train 1 at each station is all equal to $TD_1^m(1) - u_m$. Besides, there are bound constraints for opening time of initial stations in each line, which are given as follows:

$$u_m^{\min} \leq u_m \leq u_m^{\max}, \ u_m \in \mathbb{N}^+, \ m \in \mathcal{M}. \tag{3}$$

Note that the above constraints can be considered local constraints since they correspond to the timetable of each individual line. When applying a distributed optimization method, these constraints can be naturally decomposed into individual subproblems.

2.1.2. The passenger load dynamics with coupled constraints

The number of onboard passengers needs to be introduced as a state variable to explicitly describe the number of transferring passengers. Figure 1 depicts the dynamic load of train i at station k, including the inbound, outbound, and transferring passengers. The station-to-station dynamic passenger load of train i at station k on Line m is given by

$$p_i^m(k) = p_i^m(k-1) + pe_i^m(k) - ps_i^m(k), \tag{4}$$

where $p_i^m(k)$ is the onboard passengers of train i when departing from station k on Line m, $pe_i^m(k)$ and $ps_i^m(k)$ are the numbers of passengers entering and exiting train i at station k on Line m. Specifically, if station k is a non-transfer station, $pe_i^m(k)$ is simply the number of inbound passengers from the entrance of station k; otherwise, $pe_i^m(k)$ consists of arriving passengers coming from entrance and transfer passengers coming from other lines. Moreover, $b_i^m(k)$ is the passenger arrival rate on Line m at station k during train $i-1$' and i' departures. Then the number of entering passengers from the entrance of station k, which is proportional to the headway time, can be written as $(TD_i^m(k) - TD_{i-1}^m(k)) \cdot b_i^m(k)$. Thus $pe_i^m(k)$ is formulated as

$$pe_i^m(k) = (TD_i^m(k) - TD_{i-1}^m(k)) \cdot b_i^m(k) + \sum_{\substack{m' \neq m \\ m' \in \mathcal{M}}} \sum_{i'=1}^{N_{m'}} \hat{\alpha}_{mi}^{m'i'}(k) c_{mi}^{m'i'}(k), \tag{5}$$

where $\hat{\alpha}_{mi}^{m'i'}(k)$ is a binary variable indicating whether train i' on Line m' and train i on Line m are connected. We will explain this variable in detail in the next subsection. $c_{mi}^{m'i'}(k)$ is the number of passengers transferring from train i' on Line m' to train i on Line m at station k. In equation (5), the first part represents the inbound passengers from outside the metro system and the second part is the transferring passengers from other lines. In addition, the number of exiting passengers is proportional to onboard passengers:

$$ps_i^m(k) = \gamma_i^m(k) p_i^m(k-1), \tag{6}$$

where $\gamma_i^m(k)$ is the proportion factor of passenger load. Similarly, transferring passengers $c_{mi}^{m'i'}(k)$ is proportional to the exiting passengers from train i' at station k, that is,

$$c_{mi}^{m'i'}(k) = \beta_{mi}^{m'i'}(k') ps_{i'}^{m'}(k'). \tag{7}$$

The given parameters $b_i(k)$, $\gamma_i^m(k)$ and β_i^m can be estimated according to historical statistics.

Hence, by combining (4)-(7), we can rewrite the passenger load dynamic (4) as

$$\begin{aligned} p_i^m(k) =& p_i^m(k-1) + (TD_i^m(k) - TD_{i-1}^m(k)) \cdot b_i^m(k) - \gamma_i^m(k) p_i^m(k-1) \\ &+ \sum_{\substack{m' \neq m \\ m' \in \mathcal{M}}} \sum_{i'=1}^{N_{m'}} \hat{\alpha}_{mi}^{m'i'}(k) \beta_{mi}^{m'i'}(k) \gamma_{i'}^{m'}(k) p_{i'}^{m'}(k-1). \end{aligned} \tag{8}$$

Specifically, $p_i^m(0) = 0$.

Equations (1), (2) and (8) reflect the dynamics in headway and passenger load. Equation (8) implies that the passenger load results from the passenger arrival rates and train headway of connected lines, and together they form a nonlinear coupled system. The proposed model contributes to extending the model in Wong et al. (2008) with dynamic transferring passengers instead of known constant values. In addition, constraint (7) is a local constraint since it only contains variables corresponding to individual lines, while constraint (8) is a coupled constraint due to the existence of variables from different lines, i.e., p_i^m and $p_{i'}^{m'}$.

2.1.3. Transfer waiting time with coupling constraints

Wong et al. (2008) defined the transfer waiting time as the minimum possible transfer time, which is illustrated in Figure 2. In essence, only the successful transfer waiting time is counted and the others are set to zero.

To correctly depict the waiting time of passengers who transfer from train i' on Line m' to train i on Line m at station k, binary variables $\alpha_{mi}^{m'i'}(k)$ are defined in the following constraints for all $(m,\ m',\ k) \in T$:

$$C(\alpha_{mi}^{m'i'}(k) - 1) \le TD_i^m(k) - TA_{i'}^{m'}(k) - e_m^{m'}(k) \le C \cdot \alpha_{mi}^{m'i'}(k), \tag{9}$$

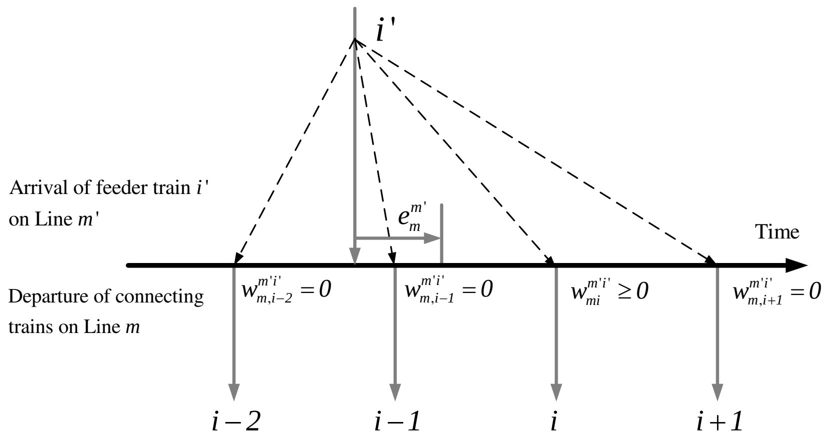

Figure 2: Illustration of transfer waiting time.

where C is a large enough positive number. $TA_{i'}^{m'}(k)+e_m^{m'}(k)$ represents the earliest arrival time of passengers from train i' of Line m' at the track of train i on Line m. Therefore $\alpha_{mi}^{m'i'}(k) = 1$ means train i' on Line m' arrives at station k early enough for passengers to catch train i on Line m; conversely, $\alpha_{mi}^{m'i'}(k) = 0$. We then define binary variable $\hat{\alpha}_{mi}^{m'i'}(k)$ as

$$\hat{\alpha}_{mi}^{m'i'}(k) = \alpha_{mi}^{m'i'}(k) - \alpha_{m,i-1}^{m'i'}(k). \tag{10}$$

Here $\hat{\alpha}_{mi}^{m'i'}(k)$ represents the actual connection of trains from connected lines. Specifically, if train i' in Line m' is not able to connect with train $i-1$ but train i in Line m, it can be derived that $\alpha_{m,i-1}^{m'i'}(k) = 0$, $\alpha_{mi}^{m'i'}(k) = 1$ and $\alpha_{m,i+1}^{m'i'}(k) = 1$. Then according to constraint (10), we can get that $\hat{\alpha}_{mi}^{m'i'}(k) = 1$ and $\hat{\alpha}_{m,i+1}^{m'i'}(k) = 0$. In other words, only the first feasible connection is considered in this study, which is consistent with the definition of transfer waiting time.

Consequently, the transfer waiting time can be computed in the following equation:

$$w_{mi}^{m'i'}(k) = [TD_i^m(k) - TA_{i'}^{m'}(k) - e_m^{m'}(k)] \cdot \hat{\alpha}_{mi}^{m'i'}(k). \tag{11}$$

According to above constraints (10) and (11), when

$$\begin{cases} TD_j^m(k) - TA_{i'}^{m'}(k) - e_m^{m'}(k) \leq 0, \ j = 1, \ldots, i-1, \\ TD_j^m(k) - TA_{i'}^{m'}(k) - e_m^{m'}(k) \geq 0, \ j = i, i+1, \ldots, N_m, \end{cases}$$

we have

$$\begin{cases} \alpha_{mj}^{m'i'}(k) = 0, \ j = 1, \ldots, i-1, \\ \alpha_{mj}^{m'i'}(k) = 1, \ j = i, i+1, \ldots, N_m, \end{cases} \quad \text{then} \quad \begin{cases} \hat{\alpha}_{mj}^{m'i'}(k) = 0, \ j \neq i, \\ \hat{\alpha}_{mj}^{m'i'}(k) = 1, \ j = i. \end{cases}$$

In this way, from train i' to train i, transfer waiting time $w_{mi}^{m'i'}(k) = TD_i^m(k) - TA_{i'}^{m'}(k) - e_m^{m'}(k) \geq 0$, and $w_{mj}^{m'i'}(k)$, $j \neq i$ are counted as 0. Different from the inequality constraints in Wong et al. (2008); Wu et al. (2015), equality (11) can accurately calculate the transfer waiting time shown in Figure 2.

When the departure and arrival times are considered dynamic variables, the connections for trains at interchange stations are unknown at the time when trains are dispatched from the initial stations. As such, the passenger transference and loads cannot be predetermined, either. Constraints (9)-(11) effectively identify the correct train connections of each train. Combined with constraint (8), the proposed model can accurately represent the transfer waiting times and passenger load dynamic for non-periodic timetables in a metro network.

There are bounds constraints for transfer waiting time in accordance with service quality, which can be written as:

$$0 \leq w_{mi}^{m'i'}(k) \leq w_{\max}^{m',m}. \tag{12}$$

In this part, similar to constraints (1)-(3), (7), the presented constraints (10) and (12) are local constraints. Constraints (9) and (12) are coupled constraints, which make the formulated model inseparable. Consensus would be a valid technique for turning the coupled constraints, which do not split due to coupled variables, into separable local constraints, which split easily (Boyd et al., 2011). Corresponding details will be introduced in Section 3.

2.2. Objective Functions

Though enhancing transfer synchronization among different lines is important, it should not be the only objective, as the waiting time of inbound passengers is also an important service performance measure. In this paper, we aim to minimize the total passenger waiting time, i.e., the linear weighted combination of the waiting time for both transferring and inbound passengers of the whole network. The specific objective function is:

$$J = \min_{\boldsymbol{TD},\boldsymbol{TA},\boldsymbol{u}} \sum_{(m,m',k)\in T} \sum_{i'=1}^{N_{m'}} \sum_{i=1}^{N_m} \theta_1 c_{mi}^{m'i'}(k) w_{mi}^{m'i'}(k) + \sum_{m=1}^{M} \sum_{i=1}^{N_m} \sum_{k=1}^{K_m} \theta_2 (TD_i^m(k) - TD_{i-1}^m(k))^2 \cdot b_i^m(k). \tag{13}$$

The first part denotes the transfer waiting time. In the second part, $b_i^m(k)(TD_i^m(k) - TD_{i-1}^m(k))$ represents the number of arrived passengers during the time interval $[TD_{i-1}^m(k), TD_i^m(k)]$. Then the waiting time of inbound passengers during the time interval equals the number of arrived passengers $b_i^m(k)(TD_i^m(k) - TD_{i-1}^m(k))$ times the average waiting time for each passenger $\frac{1}{2}(TD_i^m(k) - TD_{i-1}^m(k))$ during this time interval, i.e., $(TD_i^m(k) - TD_{i-1}^m(k))^2 \cdot \frac{b_i^m(k)}{2}$. The denominator is omitted as it can be accommodated by the weight coefficient θ_2. Note that minimizing the first term tends to enhance the connection among different lines, which may lead to a large headway for some lines. Then the second term can control the passengers waiting time to some extent. Since these two terms are the same unit, we do not normalize them but only adopt the linear weighted method. By using pre-specified weights θ_1 and θ_2, the combination of these two terms effectively achieves the desired balance between transfer waiting times and station waiting times as sought by the metro network managers.

2.3. Optimization Model

The objective function (13), together with headway dynamic equations (1), passenger load dynamic equations (8), transfer waiting time constraints (9)-(11) and bound constraints (2), (3), (12), formulate the timetable synchronization optimization problem as follows, which is referred as TSO afterward.

$$
J = \min_{\boldsymbol{TD},\boldsymbol{TA},\boldsymbol{u}} \sum_{(m,m',k)\in T} \sum_{i'=1}^{N_{m'}} \sum_{i=1}^{N_m} \theta_1 c_{mi}^{m'i'}(k) w_{mi}^{m'i'}(k) + \sum_{m=1}^{M} \sum_{i=1}^{N_m} \sum_{k=1}^{K_m} \theta_2 (TD_i^m(k) - TD_{i-1}^m(k))^2 \cdot b_i^m(k)
$$

$$
\text{s.t.} \begin{cases} TA_i^m(k+1) = TD_i^m(k) + r_i^m(k) \\ H_{\min}^m \le TD_{i+1}^m(k) - TD_i^m(k) \le H_{\max}^m \\ S_{\min}^m \le TD_i^m(k) - TA_i^m(k) \le S_{\max}^m \\ c_{mi}^{m'i'}(k) = \beta_{i'}^{m'}(k')\gamma_{i'}^{m'}(k')p_{i'}^{m'}(k'-1) \\ p_i^m(k) = p_i^m(k-1) + (TD_i^m(k) - TD_{i-1}^m(k)) \cdot b_i^m(k) - \gamma_i^m(k) p_i^m(k-1) \\ \quad + \sum_{\substack{m'\neq m \\ m'\in\mathcal{M}}} \sum_{i'=1}^{N_{m'}} \hat{\alpha}_{mi}^{m'i'}(k) c_{mi}^{m'i'}(k) \\ C(\alpha_{mi}^{m'i'}(k) - 1) \le TD_i^m(k) - TA_{i'}^{m'}(k) - e_m^{m'}(k) \le C \cdot \alpha_{mi}^{m'i'}(k) \\ \hat{\alpha}_{mi}^{m'i'}(k) = \alpha_{mi}^{m'i'}(k) - \alpha_{m,i-1}^{m'i'}(k) \\ w_{mi}^{m'i'}(k) = [TD_i^m(k) - TA_{i'}^{m'}(k) - e_m^{m'}(k)] \cdot \hat{\alpha}_{mi}^{m'i'}(k) \\ 0 \le w_{mi}^{m'i'}(k) \le w_{\max}^{m',m}, \\ u_m^{\min} \le u_m \le u_m^{\max},\ k = 1, 2, \ldots, K_m,\ i = 1, 2, \ldots, N_m. \end{cases} \tag{14}
$$

With the nonlinear constraints (8) and (11), and the nonlinear objective function, the above TSO model is essentially a large MINP, whose complexity is dependent upon the number of lines (i.e., M), trains (i.e., N_m), and stations (i.e., K_m). When considering a practical and large network, the formulation can easily become a large-scale MINP with an intractable computational burden.

By extending the NP-hard MIP problem proposed by Wong et al. (2008), the formulated TSO problem in this paper is more complicated and therefore also belongs to the NP-hard class problems. Traditional centralized optimization methods usually encounter high computational complexity and require a large computation time to solve such problems. In contrast, a distributed optimization algorithm can effectively reduce the computational burden as the overall optimization is decomposed into a series of subproblems with low dimensionality. Hence, this study constructs a distributed optimization method utilizing the ADMM framework to decompose the original optimization problem. The coupled variables are determined as arrival and departure time at transfer stations, number of transfer passengers and transfer waiting time, which are all continuous variables.

3. Distributed optimization method design

Decomposition methods are widely adopted as efficient solutions to deal with problems in large-scale networks. solutions can be obtained by addressing a series of subproblems sequentially or concurrently by exchanging information during the iterations. For instance, Li et al. (2018) used a dual decomposition technique to implement the optimal control for the large-scale train regulation problem. Kulkarni et al. (2018) decomposed the multiple depots vehicle scheduling problem for each trip using Dantzig-Wolfe decomposition. Mohebifard and Hajbabaie (2019) optimized the network-wide traffic signal based on the Benders decomposition technique. Tong et al. (2019) handled the dynamic traffic assignment problem in a complex subway network with an ADMM-based solution approach. In recent years, ADMM based decomposition approach has become popular as it can converge under relaxed conditions and make the problem decomposable, then each subproblem can be solved individually. ADMM not only has the fast convergence characteristic of the augmented Lagrangian method but also improves theoretical and practical convergence properties (Boyd et al., 2011). In this study, we adopt the ADMM-based decomposition method to handle the interconnecting constraints among different lines. Note that the original optimization problem is non-convex due to the existence of binary variables $\boldsymbol{\alpha}$ and $\hat{\boldsymbol{\alpha}}$. Then the convergence can not be guaranteed by directly applying the standard ADMM. To provide a feasible suboptimal solution, a novel two-level ADMM-based algorithm is designed to address this problem. The upper level determines the connections among trains of different lines, i.e., $\boldsymbol{\alpha}$ and $\hat{\boldsymbol{\alpha}}$, subsequently, the lower level applies standard ADMM with fixed $\boldsymbol{\alpha}$ and $\hat{\boldsymbol{\alpha}}$ to optimize the timetable.

In this section, the TSO is decomposed into a series of single-line timetabling subproblems thanks to the introduction of a global common variable. Then, the original TSO is solved iteratively via standard ADMM.

The standard ADMM procedure for addressing the proposed timetable synchronization problem is first introduced in Section 3.1. Finally, the two-level ADMM-based approach is provided in Section 3.2.

3.1. Distributed timetable synchronization optimization based on standard ADMM

We first introduce some related concepts to facilitate the implementation of ADMM. The set T is defined to identify whether line m and m' are connected by a transfer station k, if so, we have $(m, m', k) \in T$. Then line m and m' are called neighbors when $(m, m', k) \in T$. Thus, the set of neighbors of line m, which have direct interactions with line m, is defined as $\mathcal{N}_m = \{m' | (m, m', k) \in T\}$. Taking the simple three-line network shown in Figure 3(a) for illustration, line 2 is the neighbor of line 1 in the network.

Let coupled variables denote the variables that relate to a transfer station, and coupled constraints denote constraints involving coupled variables from different lines. In optimization model (14), constraints (1), (2), (7), (10) and (12) are local constraints as they are only related to the local timetable of each line. While constraints (8), (9) and (11) are coupled and not naturally separable constraints, owing to coupled variables from the neighbor lines at transfer stations such as arrival and departure time, transfer waiting time and transfer passengers. For instance, in subsystem L_1, coupled variables $TA_i^2(a)$, $TD_i^2(a)$, $w_{1,i}^{2,i'}(a)$ and $c_{1,i}^{2,i'}(a)$ of L_2 are involved in coupled constraints (8), (9) and (11).

To make optimization model (14) separable, in the proposed algorithm, we add a copy of coupled variables from neighbor subsystem m' into subsystem m. For instance, $TA^2(a)$ and $TD^2(a)$ in L_2 are duplicated in neighbor subsystem L_1 as $TA_{L_1}^2(a)$ and $TD_{L1}^2(a)$, as shown in Figure 3(b). In this way, constraints (8), (9) and (11) can relate to local coupled variables instead of coupled variables from neighbor subsystems. The consistency between the copy and original coupled variables is guaranteed by consensus constraints that make the copied and original variables equal to the global common variables, for example, $TD^1(a) = z_1,\ TD_{L_2}^1(a) = z_1$. This network can then be decomposed into three subsystems L_1, L_2 and L_3, linked by the consensus constraints and global common vectors $\boldsymbol{z_1},\ \boldsymbol{z_2}$ and $\boldsymbol{z_3}$ as shown in Figure 3(b).

Moreover, the TSO problem can be reformulated in the form of a global consensus problem, which is presented as the following proposition.

Proposition 3.1. *The TSO problem can be reformulated in the form of a global consensus problem:*

$$\min_{X_m \in \mathcal{C}_m} J = \sum_{m=1}^{M} \Phi_m(X_m) \tag{15}$$

$$s.t.\ \tilde{X}_m = E_m \boldsymbol{z},\ i = 1, 2, \ldots, N_m, \tag{16}$$

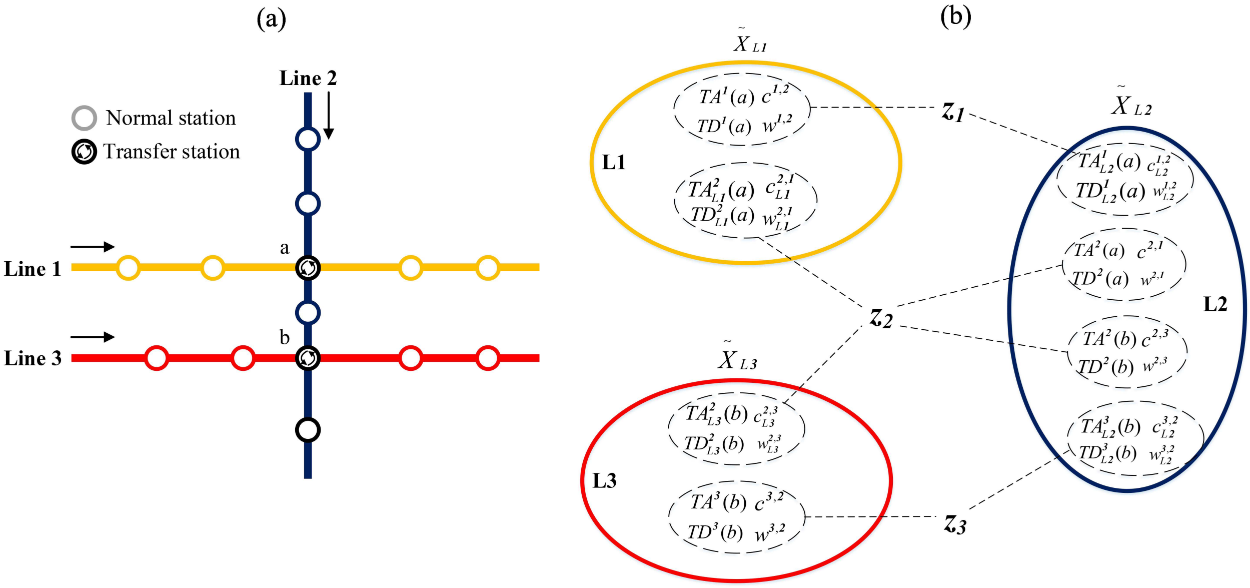

Figure 3: (a) Map of a three-line metro network; (b) Illustration for the decomposition and consensus optimization.

where the vector X_m *is the variables for all the trains in line* m*, which consists of local vector* $\bar{X}_m$ *and coupled vector* $\tilde{X}_m$*.*

Proof. Let the vector X_m be the variables for all the trains in line m, which consists of local vector $\bar{X}_m$ and coupled vector $\tilde{X}_m$, i.e., $X_m = [\bar{X}_m, \tilde{X}_m]$. Specifically, vector $\bar{X}_m$ includes station opening time, number of passengers onboard, arrival and departure time at non-transfer stations and connection indicator variables in subsystem m. Coupled vector $\tilde{X}_m$ contains transfer waiting time, number of transfer passengers, arrival and departure time at transfer stations in subsystem m, and duplicated counterparts from neighbor subsystem m'. For example, as shown in Figure 3(b), $\tilde{X}_{L_1}$ for subsystem L_1 contains $\{TA^1(a), TD^1(a), c^{1,2}, w^{1,2}\}$ originally from subsystem L_1 and duplicates $\{TA^2_{L_1}(a), TD^2_{L1}(a), c^{2,1}_{L1}, w^{2,1}_{L1}\}$ from subsystem L_2.

By introducing a vector of global common variables $\boldsymbol{z} = (\tilde{X}_m \cup \tilde{X}_{m'}, m \neq m' \in \mathcal{M})$, each of these coupled vector $\tilde{X}_m$ are a selection of the components of $\boldsymbol{z}$, i.e., $\tilde{X}_m = E_m \boldsymbol{z}$, where E_m is an incidence matrix that maps the coupled and global variables, then the coupled variables in different subsystems are linked. Thus the TSO problem can be reformulated in the form of a global consensus problem (15). □

For the distributed TSO problem, the objective (13) of the whole network is represented by the summation of objectives of individual subsystems as (15). The constraints set $\mathcal{C}_m$ contains constraints (1)-(12) for line m and its neighbors $\mathcal{N}_m$. Constraint (16) guarantees the consensus among different subsystems that duplicated variables are equal to the originals. So far, (15) and (16) present a separable consensus optimization formulation that can be solved via ADMM.

Let $\bar{z}_m = E_m \boldsymbol{z}$, the augmented Lagrangian function for problem (15) is

$$\begin{aligned} L_\rho(X_m, \boldsymbol{z}, \boldsymbol{\lambda}) &= \sum_{m=1}^{M} \left[\Phi_m(X_m) + \lambda_m(\tilde{X}_m - \bar{z}_m)^T + \frac{\rho}{2} \|\tilde{X}_m - \bar{z}_m\|_2^2 \right] \\ &= \sum_{m=1}^{M} L_{\rho,m}(X_m^{k+1}, \boldsymbol{z}^k, \boldsymbol{\lambda}^k), \end{aligned} \tag{17}$$

where $\boldsymbol{\lambda}$ are Lagrangian multipliers and ρ is the augmented Lagrangian parameter. Obviously, function (17) can be naturally separated into M independent single line train timetabling problem. Then the standard ADMM procedure is given in Algorithm 3.1.

Algorithm 3.1 ADMM-based algorithm for TSO of metro networks.

Step 1. Initialize the vector $\boldsymbol{z}$, Lagrangian multipliers $\boldsymbol{\lambda}$ and the augmented Lagrangian parameter ρ. Set $k = 1$.

Step 2. Reformulate subproblem (18) of each line using the method in Section 3.2 and solve them with the given global variable $\bar{z}_m^k$ and multipliers $\boldsymbol{\lambda}^k$.

$$\begin{aligned} X_m^{k+1} &= \arg \min_{X_m \in \mathcal{C}_m} \Phi_m(X_m) + \lambda_m^k(\tilde{X}_m - \bar{z}_m^k)^T + \frac{\rho}{2} \|\tilde{X}_m - \bar{z}_m^k\|_2^2 \\ &= \arg \min_{X_m \in \mathcal{C}_m} \Phi_m(X_m) + \lambda_m^k \tilde{X}_m^T + \frac{\rho}{2} \|\tilde{X}_m - \bar{z}_m^k\|_2^2 \end{aligned} \tag{18}$$

Step 3. Solve the z-update step (19) by collecting all the information from each line.

$$\begin{aligned} \boldsymbol{z}^{k+1} &= \arg \min_z L_\rho(X_m^{k+1}, \boldsymbol{z}, \lambda^k) \\ &= \arg \min_z \sum_{m=1}^{M} \left[\Phi_m(X_m^{k+1}) + \lambda_m^{k+1}(\tilde{X}_m^{k+1} - \bar{z}_m)^T + \frac{\rho}{2} \|\tilde{X}_m^{k+1} - \bar{z}_m\|_2^2 \right] \\ &= \arg \min_z \sum_{m=1}^{M} \left[-\lambda_m^{k+1} \bar{z}_m^T + \frac{\rho}{2} \|\tilde{X}_m^{k+1} - \bar{z}_m\|_2^2 \right] \end{aligned} \tag{19}$$

Step 4. Termination check. The iteration stops and releases the final solution if:

1) $\left[\sum_{m=1}^{M} (\tilde{X}_m^k - E_m \boldsymbol{z}^k)^2 \right]^{\frac{1}{2}} \leq \epsilon$, where ϵ is a small positive scalar.

2) the algorithm has reached the maximum iterative number $K^{\max}$.

Otherwise, go to *Step 5*.

Step 5. Update Lagrangian multipliers λ by the following equation, set $k = k + 1$ and go to *Step 2.*

$$\lambda_m^{k+1} = \lambda_m^k + \rho(\tilde{X}_m^{k+1} - \bar{z}_m^{k+1}) \tag{20}$$

In the above ADMM algorithm, *Step 2* (i.e., x_m-updates) and *Step 5* (i.e., λ_m-updates) are conducted independently for each line m, $m = 1, 2, \ldots, M$. As $\boldsymbol{z}$ is a global variable, which is known as the central collector, *Step 3* (i.e., $z-$update) is carried out centrally. In this way, the proposed timetable synchronization problem can be distributed across multiple processors. The timetable optimization problem for each line is solved by each processor, and the coupled variables of each line are updated iteratively to global common values, which is the solution to the origin problem. A flowchart is drawn in Figure 4 to illustrate the procedure of the ADMM-based method.

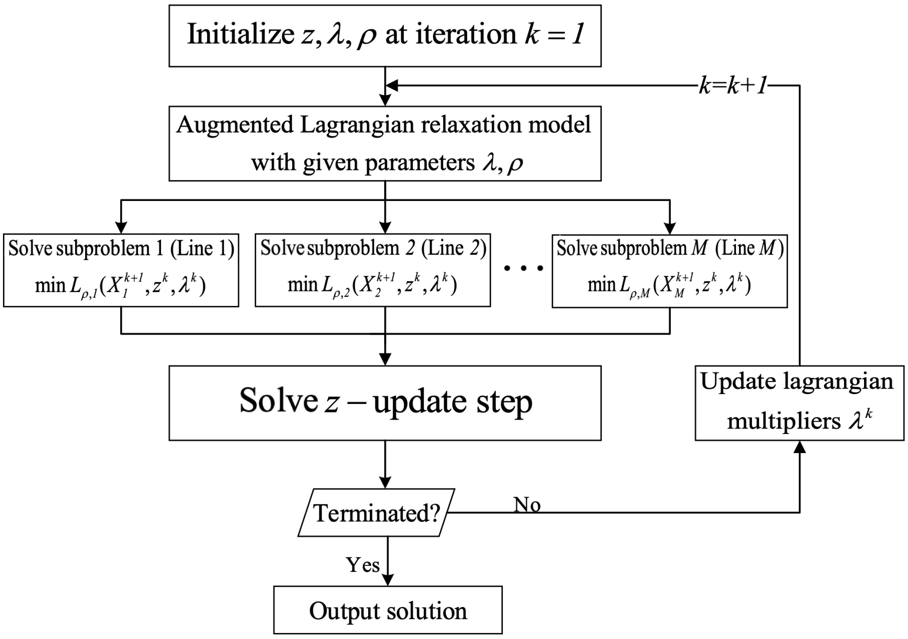

Figure 4: The illustration for the procedure of Algorithm 3.1.

3.2. ADMM based two-level approach

Due to the existence of binary variables, the original optimization model is nonconvex and may not converge using the standard ADMM approach. We design a novel two-level approach, which alternatively fixes the binary variables and the other variables, and solves these subproblems in an iterative way. The upper level first relaxes the binary variables $\boldsymbol{\alpha}$ and $\hat{\boldsymbol{\alpha}}$ as continuous values, then solves the linear relaxed problem and the timetable-fixed problem, and passes the solved binary values to the lower level. The lower level solves the binary-fixed problem and passes the solved values of timetable variables to the upper level. In accordance to the described structure of the two-level approach, **Algorithm 3.2** gives a summary of the procedure as follows.

Algorithm 3.2 Two-level ADMM approach for timetable synchronization optimization of metro networks.

Step 1. Initialization. Set $\omega = 1$, relax the binary variables $\boldsymbol{\alpha}$ and $\hat{\boldsymbol{\alpha}}$ into [0,1]. Use **Algorithm 3.1** to solve the linear relaxed problem, and output timetable variables $\boldsymbol{TD}^{\omega}$, $\boldsymbol{TA}^{\omega}$ and $\boldsymbol{u}^{\omega}$.

Step 2. Fix the timetable of each lines with given $\boldsymbol{TD}^{\omega}$, $\boldsymbol{TA}^{\omega}$ and $\boldsymbol{u}^{\omega}$. Then solve each subproblem where $\boldsymbol{\alpha}$ and $\hat{\boldsymbol{\alpha}}$ are binary variables, and obtain $\boldsymbol{\alpha}^{\omega}$ and $\hat{\boldsymbol{\alpha}}^{\omega}$ by Gurobi solver.

Step 3. Fix the transfer indicators $\boldsymbol{\alpha}^{\omega}$ and $\hat{\boldsymbol{\alpha}}^{\omega}$. Then use **Algorithm 3.1** to solve the bilinear programming problem (18) and get updated timetable variables $\boldsymbol{TD}^{\omega+1}$, $\boldsymbol{TA}^{\omega+1}$ and $\boldsymbol{u}^{\omega+1}$.

Step 4. Termination check. If $\boldsymbol{\alpha}^{\omega}$ and $\hat{\boldsymbol{\alpha}}^{\omega}$ are the same as $\boldsymbol{\alpha}^{\omega-1}$ and $\hat{\boldsymbol{\alpha}}^{\omega-1}$, then output the final solution, otherwise, set $\omega = \omega + 1$, and repeat *Step 2-Step 4*.

Figure 5 demonstrates the flowchart of the developed two-level approach. Given the fixed binary variables $\boldsymbol{\alpha}$ and $\hat{\boldsymbol{\alpha}}$, TSO constitutes a bilinear programming program solved using the standard ADMM algorithm in a decentralized fashion within finite iterations. Also, given the fixed timetable, binary variables $\boldsymbol{\alpha}$ and $\hat{\boldsymbol{\alpha}}$

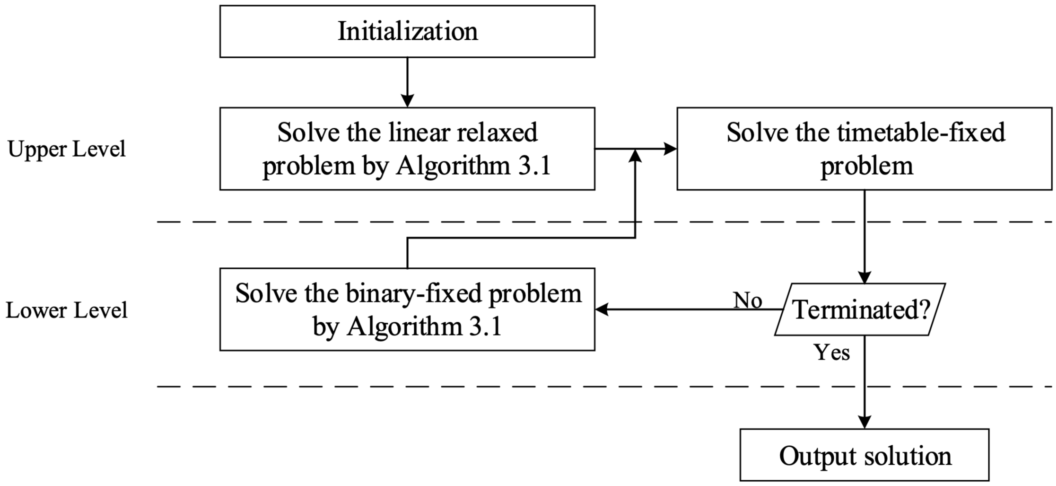

Figure 5: The flowchart of Algorithm 3.2.

are effectively solved by a solver. Therefore, the TSO problem is split into a set of single-line optimization problems which can be handled by the proposed two-level approach in parallel. In addition, under this two-level ADMM approach, the objective value decreases with iterations and finally reduces to a suboptimal solution to the network TSO.

It is worth noting that the convergence of the decomposition method is not guaranteed when dealing with complex systems and thus an exact solution should not be expected. Boyd et al. (2011) provided the theoretical proof of the convergence of the two-block ADMM under three assumptions: a) the original problem is convex; b) the unaugmented Lagrangian problem has a saddle point; c) All the subproblems have solutions. The TSO problem in this research is formulated as a multi-block minimization model with a bilinear objective function, the optimality and convergence of the ADMM-based **Algorithm 3.1** are not guaranteed, let alone that the optimal solution at each update step is not reached due to the bilinear term in the objective function. Nevertheless, even if **Algorithm 3.1** does not fully converge due to nonconvexity, its primal variables $\boldsymbol{TD}, \boldsymbol{TA}, \boldsymbol{u}$ remain feasible for local line constraints (1)-(3), (7), (10), and (12) at each iteration, as enforced by subproblems (18). If terminated before achieving full consensus, these timetable variables can be used after recomputing dependent states ($\boldsymbol{\alpha}$, $\hat{\boldsymbol{\alpha}}$, $\boldsymbol{w}$, $\boldsymbol{c}$) via constraints (8)-(11). This ensures solution feasibility for individual lines despite potential consensus gaps. In addition, we will empirically showcase the proposed algorithm's effectiveness via the numerical experiments in Section 4. Moreover, this kind of two-level approach can be extended to other large-scale transportation problems with MIP formulations.

4. Numerical examples

4.1. Network and parameter settings

We first use the example network illustrated in Figure 3(a) to showcase the effectiveness of the proposed model and distributed optimization algorithm for the TSO problem. This network can be seen as a part of the Beijing metro system, and the related parameters are chosen based on practical operational data. There are 3 lines, 4 transfer directions, and 14 stations, including 2 transfer stations and 12 non-transfer

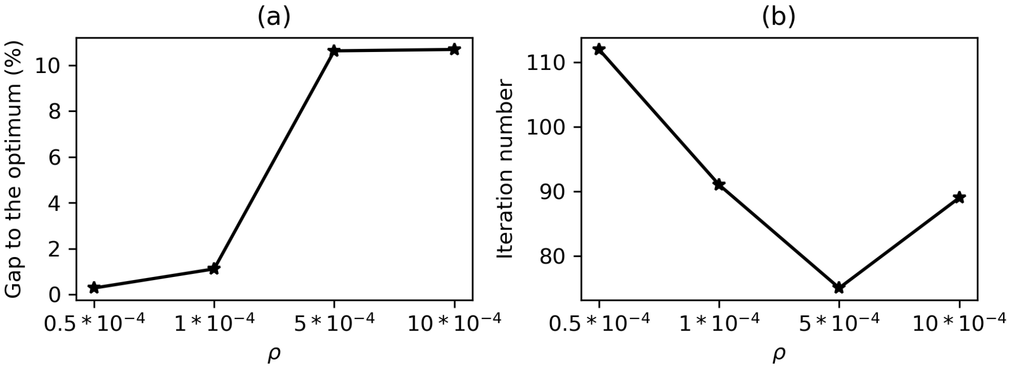

Figure 6: Impact of ρ on (a) the objective value and (b) the convergence rate for the linear relaxed timetable synchronization problem.

stations. The considered planning horizon is in a morning off-peak hour between 09:00 and 10:00. Three scenarios (S_1, S_2 and S_3) are considered, wherein 15, 21 and 30 (5, 7 and 10 trains for each line) trains are dispatched, respectively. The station-to-station running time is given by the practical operation data. The dwell time is bounded in the interval [10, 60] (unit: s) (Chu et al., 2015). The headway time for lines 1-3 are bounded in intervals [240, 390], [180, 300], [210, 360] (unit: s), respectively. The passenger demand data was recorded on a workday in 2014 from Beijing metro line 1, line 5 and line 6. The transfer walking time $e_m^{m'}$ is estimated as $30s$.

Weight coefficients $\theta_1 = 10^{-2}$ and $\theta_2 = 10^{-6}$ are adopted in objective function (13) to control the magnitude of the two terms. We generated randomly a set of 100 feasible timetables and computed transfer waiting time and inbound waiting time for each solution. For the small network experiments, the average magnitudes were 10^2 and 10^6, respectively. The same process is repeated to determine the weight coefficients used for the large network experiments. Note that alternative weights could be used depending on operator preferences (e.g., prioritizing transfers). For the parameters associated with ADMM, the initial values of $\boldsymbol{\lambda}$ are $\mathbf{0}$, and the maximum iteration number is 100. All the subproblems are implemented in MATLAB, solved by Gurobi on a desktop computer (3.5 GHz processor speed and 8-GB memory size) with the platform of Windows 10. Note that the proposed abstract mathematical model does not consider the operation constraints practically caused by signal equipment and line conditions limitations, and the parameters and factors adopted may therefore deviate from the real data, and the optimized results are not easily comparable with the practical solutions either. Thus, we showcase the efficacy of the developed approach by comparing it with the best solutions found by solvers in Section 4.2 and simulations in Section 4.3, respectively.

4.2. Computational results

Proper selection of the penalty parameter ρ is crucial to the performance of the ADMM algorithm. As pointed out by Boyd et al. (2011), larger ρ would reduce violations of primal feasibility while smaller ρ would reduce the dual residual, which may lead to a larger primal residual. In this section, we set ρ as different values from $0.5 * 10^{-4}$ to $10 * 10^{-4}$ to test the algorithm performance with ϵ of 10^{-2} under scenarios S_3. As shown in Figure 6, larger values of ρ require relatively fewer iterations to satisfy the stopping criteria, but at the expense of enlarging the gap of the optimized objective value to the optimum. Taking into account the balance between computational efficiency and solution quality comprehensively, $\rho = 1 * 10^{-4}$ is chosen in

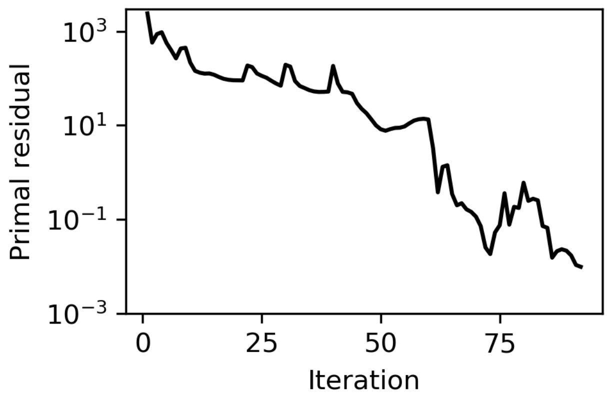

Figure 7: Primal residual evolution of the distributed ADMM algorithm.

this experiment. Figure 7 plots the convergence evolution of the distributed ADMM algorithm for the linear relaxed timetable synchronization problem, where the primal residual is reduced to predefined convergence tolerance 10^{-2} after 91 iterations, indicating that the timetable optimized by the three lines coincided. In a similar way, the lower-level subproblem of the two-level approach also has good convergence that the primal residual reaches 10^{-2} in 100 iterations.

We further measure the performance of the proposed method by comparing the solutions obtained by the two-level ADMM-based approach and the Gurobi solver, respectively. Table 1 lists the results of the Gurobi solver and the two-level approach, in terms of objective values, CPU time and gaps, for all three scenarios. In the column of CPU time, we report the total CPU time and the time to find the optimal solution in the parenthesis. Note that, in Table 1, the gap for the Gurobi solver represents the relative gap between the best solution and the best lower bound found by the solver when solving the original MINP formulation, and the gap for the two-level ADMM approach represents the gap between the obtained solution and the same lower bound provided by the solver. It can be found that, though the Gurobi solver can find the optimal solution to the complex optimization problem with a simple network, the CPU time exhibits exponential growth as the number of trains increases, despite the optimality-proof overhead. The two-level ADMM finds only sub-optimal solutions for all three scenarios, due to the fact that each sub-problems cannot be efficiently solved to the optimum. Nevertheless, the gaps ($< 1.5\%$) to the optimal solutions are acceptable, and there is a significant advantage in CPU times for scenarios with a larger number of trains. It could be expected that this optimization problem would be intractable for general solvers in a large-scale network, which can be well handled by the distributed ADMM algorithm.

Table 1: Performance comparison of Gurobi and two-level ADMM approach

Scenarios	Method	Objectives	T_1	T_2	CPU time (s)	Gap
S_1 (15 trains)	Gurobi	1.58	3.45	$1.54 \cdot 10^6$	420 (360)	0
	Two-Level ADMM	1.60	4.06	$1.56 \cdot 10^6$	102	1.17%
S_2 (21 trains)	Gurobi	2.10	3.45	$2.06 \cdot 10^6$	5730 (3500)	0
	Two-Level ADMM	2.13	6.22	$2.06 \cdot 10^6$	276	1.33%
S_3 (30 trains)	Gurobi	2.88	3.45	$2.84 \cdot 10^6$	41020 (4100)	0
	Two-Level ADMM	2.91	6.40	$2.84 \cdot 10^6$	337	1.11%

Next, we test the influence of the initialization of vector $\boldsymbol{z}$ under scenario S_3. Two kinds of initialization are conducted in this experiment: 1) initialize vector $\boldsymbol{z}$ by running *Step 2* and *Step 3* with $\rho = 0$, as stated in **Algorithm 3.1**; 2) initialize $\boldsymbol{z}$ as a zero vector, i.e., $\boldsymbol{z} = \mathbf{0}$. As expected, both initialization methods result in the same objective value, but a different number of iterations and computational time, and the later method converges faster in this experiment. This implies that the initialization of vector $\boldsymbol{z}$ will affect the solution process but not the quality of the proposed algorithm. We defer the investigation on the appropriate initialization method to future research.

Specifically for scenario S_3, it is noteworthy that all the 20 train services on Line 2 and Line 3 are successfully connected to train services on Line 1 and Line 2 at station a and b, respectively. For Line 1, only the passengers on the first 9 train services can transfer to Line 2. Additionally, all the transfer waiting time is 0 except that the transfer waiting time from train service 1 on Line 1 to train service 1 on Line 2 is 8 seconds.

4.3. Experiment for a large scale network

We apply the developed modeling framework to the real case of the Beijing metro system. Figure 8 illustrates the core portion of the Beijing metro system, including 12 lines, 28 transfer stations, 53 transfer directions, and 73 normal stations. The timetable period considered in this experiment starts at 09:00 and ends at 10:00 in the morning off-peak hour, during which 120 trains (10 trains on each line) are operated. The minimum headway times are 150s for lines 4, 14-16 and 180s for the others. The maximum headway time of all the lines is 360s. In addition, the passenger demand data was recorded on a workday in 2014. In this experiment, we also consider one direction of each line represented by the arrows in Figure 8. The transfer walking time between any two connected lines is 30s. The opening time of each line u_m is restricted from 0 to 600s. The weight coefficients set as $\theta_1 = 10^{-4}$ and $\theta_2 = 10^{-6}$, respectively. The other parameters are the same as those in Section 4.1.

The Gurobi solver fails to find any feasible solution to this large-scale problem due to the quadratic (and bilinear) terms in the objective function, thus the sub-optimal solution obtained by the proposed two-level ADMM approach is compared to the best solution found by a Monte Carlo method, containing 10^4 samples. In addition to the objective values, we introduce another metric, average transfer time, which represents the ratio of the sum of the transfer waiting time to the total successful transfer passengers, i.e., $T_{Avg} = \frac{\sum_{(m,m',k)\in T} \sum_{i'=1}^{N_{m'}} \sum_{i=1}^{N_m} c_{mi}^{m'i'}(k) w_{mi}^{m'i'}(k)}{\sum_{(m,m',k)\in T} \sum_{i'=1}^{N_{m'}} \sum_{i=1}^{N_m} c_{mi}^{m'i'}(k) \hat{\alpha}_{mi}^{m'i'}(k)}$.

The results of the Monte Carlo solution and the optimized solution are summarized in Table 2, including two objective values, average transfer time, and the number of connections. Results in Table 2 suggest that

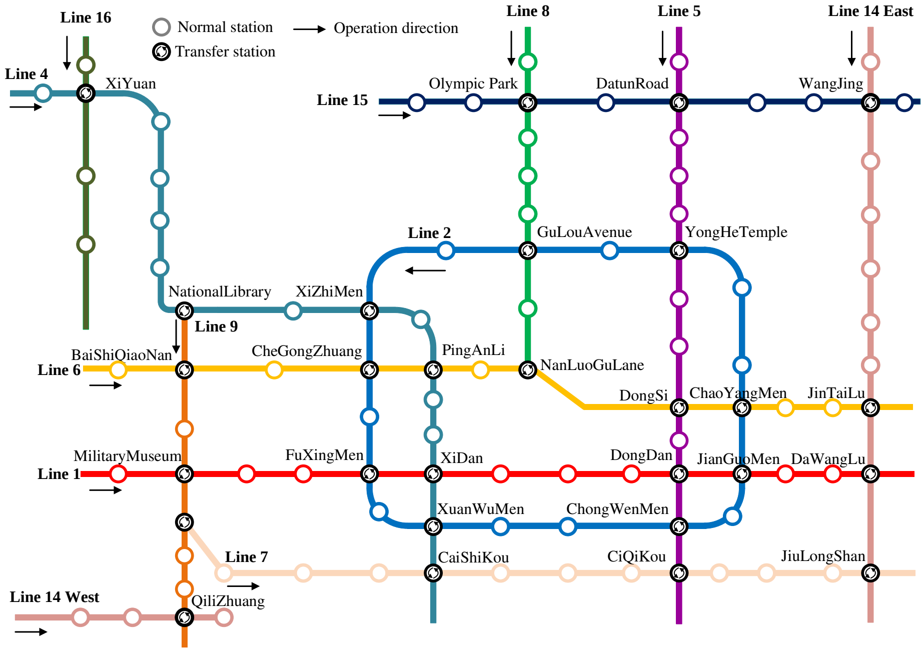

Figure 8: Map of Beijing metro network.

compared to the best solution of the Monte Carlo method, the optimized solution has a 30.3% lower total transfer waiting time and 52.0% lower total passenger waiting time, with an overall 33.6% lower objective value. A reduction in average transfer waiting time is observed (from 202.21s to 181.79s, or 10.1%), while the connected train services are reduced by 17. In addition, there is a negative correlation between the average transfer waiting time and the number of transfer passengers. In other words, the average transfer waiting time of one station is shorter when there are more transferring passengers. For example, the average number of transfer passengers from Line 1 to Line 5 at DongDan station is about 10, while there are about 31 passengers transferring from Line 6 to Line 5 at DongSi station. T_{Avg} at DongDan is about 118s while at DongSi it is only about 33s. This indicates that this model can realize relatively balanced differentiation for the timetable synchronization problem to some extent. Moreover, the transfer waiting time of the four stations crossed by Line 1, Line 2 and Line 6, including 8 directions, is illustrated in Figure 9, in which the solution of Monte Carlo method is colored in orange, while the one of the two-level ADMM approach is colored in blue, and the unsuccessful transfers are represented by gray bars. It can be found that the transfer waiting time for most of the transfer directions is apparently reduced.

Table 2: Optimized results comparison

Objectives	Monte Carlo	Optimized	Gap
Transfer waiting time	$2.44 \cdot 10^6$	$1.70 \cdot 10^6$	30.3%
Passenger waiting time	$4.15 \cdot 10^7$	$1.99 \cdot 10^7$	52.0%
Average transfer time	202.21	181.79	10.1%
Connections	444	427	-3.8%

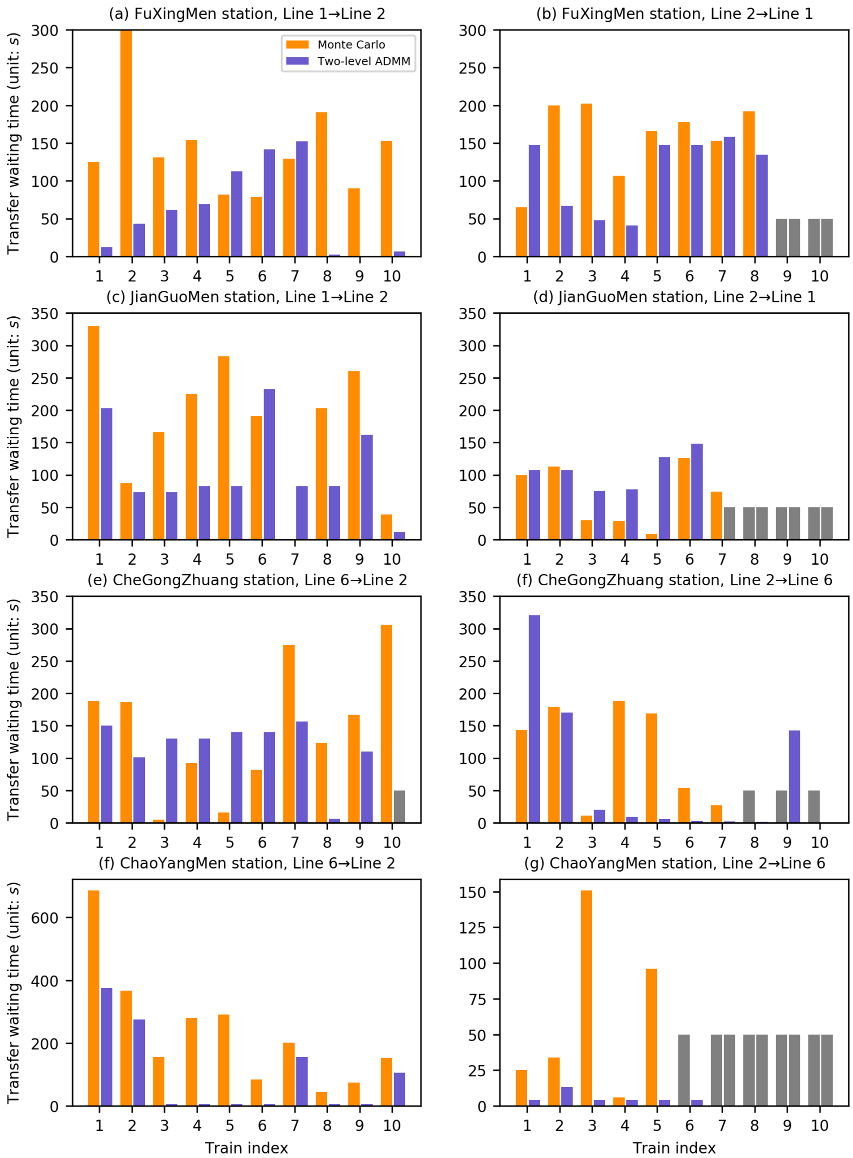

Figure 9: Illustration of the transfer waiting time between two connected lines.

We further validated robustness by conducting three additional tests where 60 (around 60% of all stations) randomly selected stations had their passenger arrival rates adjusted within ±20%. It was found that ADMM reduced transfer waiting time, passenger waiting time and average transfer time compared to Monte Carlo across all experiments, similar to the instance given above. These results suggest ADMMs advantages are not instance-specific.

5. Conclusions

In this paper, we studied the TSO problem in a complex metro network, which aims to decrease the combined waiting time for both transferring and inbound passengers in the whole network. By considering the dynamic load of passengers and introducing the definition of transfer waiting time, a nonlinear optimization model was formulated. In particular, we introduced coupled variables and global common variables based on the consensus technique to handle the inseparability of the coupled constraints. To cope with the nonconvexity issue, a two-level ADMM-based approach was particularly developed to split the network problem into a set of subproblems corresponding to the number of operating lines. This method can handle the complexity of the coupled original problem quickly in a parallel manner, which effectively reduces the computational burden. In addition, the proposed two-level ADMM approach provides a new way for addressing the large-scale transportation problem which is formulated as MIP.

A series of numerical experiments using the Beijing metro system was implemented to show the efficacy of the developed model and method. Results implied that the choices of parameter ρ would have an influence on the convergence rate and solution quality. Specifically, a large ρ will lead to a worse solution but might converge faster, and the value of ρ is determined by trial and error. Though the optimality of the proposed two-level ADMM algorithm is not guaranteed, the numerical results showcased the efficacy of the developed approach empirically, that it is able to obtain sufficiently good solutions (with gaps to the optimum smaller than 1.5%) at a relatively faster speed for the small network. Besides, the results of the large-scale numerical experiment demonstrated that the ADMM-based algorithm is also able to effectively solve the optimization problem to an acceptable quality, with smaller transfer waiting time, passenger waiting time, and average transfer waiting time compared to the Monte Carlo method. Notably, there is a negative correlation between the average transfer waiting time and the number of transfer passengers.

There are several possible avenues for future work. In the current work, the mathematical evaluation of the solution quality is absent. It is worth investigating methods to provide a (lower) bound to the optimum for general cases. Second, we propose to integrate dynamic transfer walking times as functions of passenger volume and station layout, calibrated via pedestrian simulation tools. Additionally, data-driven dwell time models could explicitly link boarding/alighting rates to dwell time adjustments, further bridging macroscopic coordination and microscopic passenger dynamics. This will complicate the TSO problem due to the increased complexity in the objective function and coupling constraints, which could be investigated in our future work.

Acknowledgements

This work was supported by the National Natural Science Foundation of China (Nos. 72431002, 72288101), and by the Beijing Natural Science Foundation, China (No. L221005).

Reference

Al Islam, S. B., Hajbabaie, A., 2017. Distributed coordinated signal timing optimization in connected transportation networks. Transportation Research Part C: Emerging Technologies 80, 272–285.

Assis, W. O., Milani, B. E., 2004. Generation of optimal schedules for metro lines using model predictive control. Automatica 40 (8), 1397–1404.

Blanco, V., Conde, E., Hinojosa, Y., Puerto, J., 2020. An optimization model for line planning and timetabling in automated urban metro subway networks. A case study. Omega, 92, 102165.

Bärmann, A., Martin, A., Schülldorf, H., 2017. A decomposition method for multiperiod railway network expansion–with a case study for germany. Transportation Science 51 (4), 1102–1121.

Bemporad, A., Morari, M., 1999. Control of systems integrating logic, dynamics, and constraints. Automatica 35 (3), 407–427.

Boyd, S., Parikh, N., Chu, E., 2011. Distributed optimization and statistical learning via the alternating direction method of multipliers. Now Publishers Inc.

Cacchiani, V., Huisman, D., Kidd, M., Kroon, L., Toth, P., Veelenturf, L., Wagenaar, J., 2014. An overview of recovery models and algorithms for real-time railway rescheduling. Transportation Research Part B: Methodological 63, 15–37.

Ceder, A. A., Hassold, S., Dano, B., 2013. Approaching even-load and even-headway transit timetables using different bus sizes. Public Transport 5 (3), 193–217.

Chu, W. J., Zhang, X. C., Chen, J. H., Xu, B. 2015. An ELMBased Approach for Estimating Train Dwell Time in Urban Rail Traffic. Mathematical Problems in Engineering, 2015(1), 473432.

Corman, F., D'Ariano, A., Marra, A. D., Pacciarelli, D., Samà, M., 2017. Integrating train scheduling and delay management in real-time railway traffic control. Transportation Research Part E: Logistics and Transportation Review 105, 213–239.

Corman, F., D'Ariano, A., Pacciarelli, D., Pranzo, M., 2012. Optimal inter-area coordination of train rescheduling decisions. Transportation Research Part E: Logistics and Transportation Review 48 (1), 71–88.

Corman, F., D'Ariano, A., Pacciarelli, D., Pranzo, M., 2014. Dispatching and coordination in multi-area railway traffic management. Computers & Operations Research 44, 146–160.

D'Ariano, A., Corman, F., Pacciarelli, D., Pranzo, M., 2008. Reordering and local rerouting strategies to manage train traffic in real time. Transportation Science 42 (4), 405–419.

Fonseca, J. P., van der Hurk, E., Roberti, R., Larsen, A., 2018. A matheuristic for transfer synchronization through integrated timetabling and vehicle scheduling. Transportation Research Part B: Methodological 109, 128–149.

Gkiotsalitis, K., Cats, O., Liu, T., 2023. A review of public transport transfer synchronisation at the real-time control phase. Transport Reviews 43 (1), 88–107.

Guo, H., Bai, Y., Hu, Q., Zhuang, H., Feng, X., 2021. Optimization on metro timetable considering train capacity and passenger demand from intercity railways. Smart and Resilient Transportation, 3 (1), 66–77.

Guo, X., Sun, H., Wu, J., Jin, J., Zhou, J., Gao, Z., 2017. Multiperiod-based timetable optimization for metro transit networks. Transportation Research Part B: Methodological 96, 46–67.

Guo, X., Wu, J., Sun, H., Liu, R., Gao, Z., 2016. Timetable coordination of first trains in urban railway network: a case study of beijing. Applied Mathematical Modelling 40 (17-18), 8048–8066.

Guo, X., Wu, J., Sun, H., Yang, X., Jin, J., Wang, D. Z., 2020. Scheduling synchronization in urban rail transit networks: Trade-offs between transfer passenger and last train operation. Transportation Research Part A: Policy and Practice, 138, 463-490.

Jin, B., Feng, X., Wang, Q., Sun, P., 2022. Real-time train regulation method for metro lines with substation peak power reduction. Computers & Industrial Engineering 168, 108113.

Kang, L., Meng, Q., 2017. Two-phase decomposition method for the last train departure time choice in subway networks. Transportation Research Part B: Methodological 104, 568–582.

Kang, L., Wu, J., Sun, H., Zhu, X., Gao, Z., 2015a. A case study on the coordination of last trains for the beijing subway network. Transportation Research Part B: Methodological 72, 112–127.

Kang, L., Wu, J., Sun, H., Zhu, X., Wang, B., 2015b. A practical model for last train rescheduling with train delay in urban railway transit networks. Omega 50, 29–42.

Kang, L., Zhu, X., Sun, H., Puchinger, J., Ruthmair, M., Hu, B., 2016. Modeling the first train timetabling problem with minimal missed trains and synchronization time differences in subway networks. Transportation Research Part B: Methodological 93, 17–36.

Kulkarni, S., Krishnamoorthy, M., Ranade, A., Ernst, A. T., Patil, R., 2018. A new formulation and a column generation-based heuristic for the multiple depot vehicle scheduling problem. Transportation Research Part B: Methodological 118, 457–487.

Kwan, C. M., Chang, C. S., 2008. Timetable synchronization of mass rapid transit system using multiobjective evolutionary approach. IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews) 38 (5), 636–648.

Le, T., Vu, H. L., Walton, N., Hoogendoorn, S. P., Kovács, P., Queija, R. N., 2017. Utility optimization framework for a distributed traffic control of urban road networks. Transportation Research Part B: Methodological 105, 539–558.

Li, S., Dessouky, M. M., Yang, L., Gao, Z., 2017. Joint optimal train regulation and passenger flow control strategy for high-frequency metro lines. Transportation Research Part B: Methodological 99, 113–137.

Li, S., Yang, L., Gao, Z., 2020. Distributed optimal control for multiple high-speed train movement: An alternating direction method of multipliers. Automatica 112, 108646.

Li, S., Zhou, X., Yang, L., Gao, Z., 2018. Automatic train regulation of complex metro networks with transfer coordination constraints: A distributed optimal control framework. Transportation Research Part B: Methodological 117, 228–253.

McCormick, G. P., 1976. Computability of global solutions to factorable nonconvex programs: Part I–Convex underestimating problems. Mathematical programming 10 (1), 147–175.

Mirheli, A., Tajalli, M., Hajibabai, L., Hajbabaie, A., 2019. A consensus-based distributed trajectory control in a signal-free intersection. Transportation Research Part C: Emerging Technologies 100, 161–176.

Mohebifard, R., Hajbabaie, A., 2019. Optimal network-level traffic signal control: A benders decomposition-based solution algorithm. Transportation Research Part B: Methodological 121, 252–274.

Niu, H., Tian, X., Zhou, X., 2015. Demand-driven train schedule synchronization for high-speed rail lines. IEEE Transactions on Intelligent Transportation Systems 16 (5), 2642–2652.

Nguyen, H. T., Chow, A. H., Ying, C. S., 2021. Pareto routing and scheduling of dynamic urban rail transit services with multi-objective cross entropy method. Transportation Research Part E: Logistics and Transportation Review, 156, 102544.

Schmaranzer, D., Braune, R., Doerner, K. F., 2020. Population-based simulation optimization for urban mass rapid transit networks. Flexible Services and Manufacturing Journal, 32 (4), 767–805.

Schröder, M., Solchenbach, I., 2006. Optimization of transfer quality in regional public transit. Tech. Rep. 84, Fraunhofer (ITWM).

Shafahi, Y., Khani, A., 2010. A practical model for transfer optimization in a transit network: Model formulations and solutions. Transportation Research Part A: Policy and Practice 44 (6), 377–389.

Tajalli, M., Hajbabaie, A., 2018. Distributed optimization and coordination algorithms for dynamic speed optimization of connected and autonomous vehicles in urban street networks. Transportation Research Part C: Emerging Technologies 95, 497–515.

Tong, L., Pan, Y., Shang, P., Guo, J., Xian, K., Zhou, X., 2019. Open-source public transportation mobility simulation engine DTALite-s: A discretized space–time network-based modeling framework for bridging multi-agent simulation and optimization. Urban Rail Transit 5 (1), 1–16.

Wang, Y., D'Ariano, A., Yin, J., Meng, L., Tang, T., Ning, B., 2018. Passenger demand oriented train scheduling and rolling stock circulation planning for an urban rail transit line. Transportation Research Part B: Methodological 118, 193–227.

Wang, Y., Tang, T., Ning, B., Meng, L., 2017. Integrated optimization of regular train schedule and train circulation plan for urban rail transit lines. Transportation Research Part E: Logistics and Transportation Review 105, 83–104.

Wang, Y., Tang, T., Ning, B., van den Boom, T. J., De Schutter, B., 2015. Passenger-demands-oriented train scheduling for an urban rail transit network. Transportation Research Part C: Emerging Technologies 60, 1–23.

Wang, Z., Ong, C. J., 2017. Distributed model predictive control of linear discrete-time systems with local and global constraints. Automatica 81, 184–195.

Wong, R. C., Yuen, T. W., Fung, K. W., Leung, J. M., 2008. Optimizing timetable synchronization for rail mass transit. Transportation Science 42 (1), 57–69.

Wu, J., Liu, M., Sun, H., Li, T., Gao, Z., Wang, D. Z., 2015. Equity-based timetable synchronization optimization in urban subway network. Transportation Research Part C: Emerging Technologies 51, 1–18.

Wu, W., Liu, R., Jin, W., Ma, C., 2019. Stochastic bus schedule coordination considering demand assignment and rerouting of passengers. Transportation Research Part B: Methodological 121, 275–303.

Ye, H., Liu, R., 2016. A multiphase optimal control method for multi-train control and scheduling on railway lines. Transportation Research Part B: Methodological 93, 377–393.

Ye, H., Liu, R., 2017. Nonlinear programming methods based on closed-form expressions for optimal train control. Transportation Research Part C: Emerging Technologies 82, 102–123.

Yin, J., Wang, M., DAriano, A., Zhang, J., Yang, L., 2023. Synchronization of train timetables in an urban rail network: A bi-objective optimization approach. Transportation Research Part E: Logistics and Transportation Review, 174, 103142.

Yin, J., Yang, L., Tang, T., Gao, Z., Ran, B., 2017. Dynamic passenger demand oriented metro train scheduling with energy-efficiency and waiting time minimization: Mixed-integer linear programming approaches. Transportation Research Part B: Methodological 97, 182–213.

Yuan, Y., Li, S., Yang, L., Gao, Z., 2022. Real-time optimization of train regulation and passenger flow control for urban rail transit network under frequent disturbances. Transportation Research Part E: Logistics and Transportation Review 168, 102942.

Zhang, H., Li, S., Yang, L., 2019. Real-time optimal train regulation design for metro lines with energy-saving. Computers & Industrial Engineering, 127, 1282–1296.

Zhao, J., Bukkapatnam, S., Dessouky, M. M., 2003. Distributed architecture for real-time coordination of bus holding in transit networks. IEEE Transactions on Intelligent Transportation Systems 4 (1), 43–51.

Zhou, X., Tong, L., Mahmoudi, M., Zhuge, L., Yao, Y., Zhang, Y., Shang, P., Liu, J., Shi, T., 2018. Open-source vrplite package for vehicle routing with pickup and delivery: a path finding engine for scheduled transportation systems. Urban Rail Transit 4 (2), 68–85.

Zhou, Y., He, S., Wang, X., Wang, P., Chen, Y., Luo, M., 2023. Optimization of Heterogeneous Passenger Subway Transfer Timetable Considering Social Equity. Urban Rail Transit 9, 246–265.