Chapter 2
Background: Train Operations and Scheduling

Abstract In this chapter, background material and literature review on the operation of trains and on urban rail train scheduling are presented. In Sect. 2.1, the operation of trains is introduced, where the automatic train operation (ATO) system is explained in detail. In addition, a brief introduction to fixed block signaling systems and moving block signaling systems is also given. An overview of optimal control approaches for the trajectory planning of a single train and multiple trains is provided in Sect. 2.2. Section 2.3 introduces the urban rail transit scheduling problem is introduced. This chapter concludes with a short summary in Sect. 2.4.

2.1 Operation of Trains

Nowadays, several dedicated high-speed railway lines and urban rail transit systems with short headways are operated with a high degree of automation [1]. This requires advanced train control systems to fulfill safety and operational requirements, such as the European train control system and communication-based train control systems, which include equipment on board of trains as well as in control centers [2]. Advanced train control systems enable the energy-efficient driving of trains, which becomes more and more important because of the operation costs and environmental concerns [3].

The automatic train operation (ATO) system of an advanced train control system drives the train according to a predefined train trajectory (i.e., a speed profile) [4] to ensure punctuality and energy saving. In addition, signaling systems in train control systems is important for running safety of trains. In this section, we first give a brief introduction to ATO systems and then provide a short introduction to the principle of signaling systems.

2.1.1 Automatic Train Operation

With the development of modern railway systems, automatic train control (ATC) systems have become vital equipment that ensures the running safety, shortens the
train headways, and improves the quality of train operations [4]. An advanced automatic train control system could consist of an automatic train protection (ATP) system, an automatic train supervision (ATS) system, and an ATO system as shown in Fig. 2.1 [5]. The onboard ATP system is responsible for supervising the train speed according to the safety speed profile and for applying an appropriate braking force if necessary. In addition, the onboard ATP system also communicates with the wayside ATP system to exchange information (e.g., temporary speed limits and the limits of movement authority (i.e., the maximum position that a train is allowed to move to)) to guarantee the safety of the operations of trains. The ATS system acts as an interface between the operator and the railway system, managing trains according to the specific regulation criteria. The ATO system controls the traction and braking force to keep the train speed under the speed limit established by the ATP system. The ATO system can be used to facilitate the driver or to operate the train in a fully automatic mode; it thus plays a key role in ensuring accurate stopping, operation punctuality, energy saving, and riding comfort [4].

An onboard ATO system consists of two levels of control actions, as conceptually illustrated in Fig. 2.2. The higher level optimizes the optimal speed-position
reference trajectory for the operation of the train, where the line resistance, speed limits, maximum traction and braking forces, etc. are taken into account. The low-level control is used to make the train track the preplanned reference trajectory via certain control methods (such as PID control, model predictive control, and robust control). The traction or braking control commands are implemented to the train and information on e.g. the speed and position of the train is collected by the sensors and transferred to the ATO system in real time.

The driving performance including punctuality, energy consumption, etc. strongly depends on the optimal reference trajectory both when the train is partly or fully controlled by the ATO system. In addition, there exist several driver assistance systems to enhance the driving performance of the drivers, such as the FreightMiser, Metromiser, and the driving style manager. The FreightMiser and Metromiser systems [6] were developed by the scheduling and control group of the University of South Australia in order to calculate the optimal reference trajectory and to give advices to the drivers of long-haul trains and suburban trains respectively. That group mainly focused on minimizing the energy consumption through Pontryagin’s principle [7]. The driving style manager [8] developed by Bombardier implements dynamic programming to calculate energy-efficient train trajectories, which are then displayed to the train driver. Whenever the train stops at a station, the driving style manager calculates the optimal trajectory to the subsequent station using real-time information. An energy-efficient driving module is integrated in the Trainguard MT communications-based train control (CBTC) system of Siemens, where a simulation-based approach is applied to obtain the energy-optimal trajectories for trains [9]. The interested reader is referred to [10] for more information on the implemented systems.

ATO systems and driver assistance systems are able to take advantage of a pre-computed train speed trajectory. However, if the operational conditions change, the ATO system will calculate an updated optimal trajectory. Therefore, it is important to design efficient algorithms to find the optimal speed-position reference trajectory. In the literature, various algorithms have been developed to optimize the speed trajectory for trains and these algorithms will be reviewed in Sect. 2.2.

2.1.2 Principles of Signaling Systems

Block signaling is used to maintain a safe distance between successive trains on the same track. There are two main types of signaling systems, namely fixed block signaling systems and moving block signaling systems. The main principles of those two signaling systems are presented next.

2.1.2.1 Fixed Block Signaling Systems

Fixed block signaling (FBS) systems are commonly used in railway operation systems nowadays [11]. In FBS systems, a track is divided into blocks, the length of which depends on the maximum train speed, the worst-case braking rate, and the
number of signal aspects, such as a green, yellow, or red. Each block is exclusively occupied by only one train and the presence of a train within a block is usually detected by the track circuits [12]. Furthermore, blocks are protected by wayside signals (i.e., signals next to the track) or cab signals (i.e., visual signals on board of trains). Wayside signals are still typical in railways, however, cab signals are used more and more, in particular on high-speed lines where wayside signals cannot be watched clearly by drivers because of the high speed. There are one-block signaling and multiple-block signaling in FBS systems [11]. In one-block signaling, the indication of the block signal depends only on the state of the block section after the signal and every block signal must have a distant signal, which is supposed to provide the required approach information. In multiple-block signaling systems, the indication of a block signal depends on the state of two or more subsequent block sections.

A simple example is a two-block signaling system with three aspects, i.e. red, yellow, and green, and which is also called a three-aspect signaling system. Such a three-aspect signaling system on a line equipped with an ATP system is shown as Fig. 2.3. Each block carries an electronic speed code through its track circuit. The speed code data consists of two parts: the authorized-speed code for this block and the target-speed code for the next block. The speed code data is coded by the electronic equipment controlling the track circuitry and is transmitted via tracks. This speed code data is then picked up by antennae on board of the train. If a train tries to enter a zero speed block or an occupied block, or if it enters a section at a speed higher than that authorized by the speed code, the onboard electronics will trigger an emergency brake.

### 2.1.2.2 Moving Block Signaling Systems

With the increasing operational density in railway systems, railway systems with an FBS system are often suffering from a shortage in transportation capacity. Even though the line capacity of an FBS system can be increased using shorter block lengths, the installation and maintenance cost of the signaling and track equipment may not be justified by the increased capacity. Consequently, moving block signaling (MBS) systems have been proposed to achieve a higher performance.

In an MBS system, the blocks are defined as dynamic safe zones around each train. Regular communication between trains and local traffic centers is needed for knowing the exact locations and speeds of all trains in the area controlled by the local traffic
2.1 Operation of Trains

Fig. 2.4 The principle of a pure MBS system

center at any given time. Therefore, compared to an FBS system, an MBS system allows trains to run closer together, thus increasing the transport capacity. The local traffic center computes the so-called limit of movement authority for every train in the area it controls and makes sure that each train will be running at a safe distance with respect to other trains (cf. Fig. 2.4). More specifically, the limit of movement authority represents the maximum position that a train is allowed to move to and it is determined by the tail of the preceding train with a safety margin included. In addition, the limit of movement authority of the following train moves forward continuously as the leading train travels. In the literature, four MBS schemes [13] have been discussed: moving space block signaling, moving time block signaling, pure MBS, and relative MBS. Takeuchi et al. [12] evaluated the first three schemes and compared them with the FBS scheme based on two basic criteria, viz., steady-state performance and perturbed performance. It is concluded that the pure MBS scheme gives the best performance. In addition, Takeuchi et al. [12] stated that the concept of the relative MBS has never been accepted for regular rail traffic even though it is routinely accepted for road traffic. Therefore, we will mainly consider the pure MBS scheme later on in this book. However, the proposed approaches can be extended to other MBS schemes too. Moreover, the pure MBS scheme is the basis of all systems currently implemented in practice [12].

In a pure MBS system, the minimum distance between two successive trains is basically the sum of the instantaneous braking distance required by the following train and a safety margin (which is introduced to avoid collisions even if the leading train comes to a sudden halt) as shown in Fig. 2.4. However, the minimum distance between trains in practice should also take the train length and the running distance during the reaction time of the drivers or ATC systems into account.

2.2 Optimal Trajectory Planning of Trains

In this section, we first give a literature review on the optimal trajectory planning of a single train and then the state-of-the-art on the trajectory planning of multiple trains with signaling constraints is reviewed.
2.2.1 Optimal Trajectory Planning of a Single Train

The research on optimal trajectory planning for a single train started in the 1960s. A simplified train optimal control problem was studied by Ichikawa [14], who solved the problem using Pontryagin’s principle. Later on, many researchers explored this optimal control problem by applying various methods, since it has significant effects for energy saving, punctuality, and riding comfort. These methods can be grouped into two main categories [8], viz., analytical solution and numerical optimization.

The aim of this section is to give an overview of the research on optimal trajectory planning. Thereby, the research reported in literature will be reviewed using these two categories.

- Analytical solution
  The train is usually modeled as a point mass in the optimal control problem. According to whether the traction and braking force is continuous or discrete, there are two kinds of models, i.e. continuous-input models and discrete-input models. The research on discrete-input models is mainly done by the SCG group of the University of South Australia [6, 15]. A type of diesel-electric locomotive is considered, the throttle of which can take only on a finite number of positions. Each position determines a constant level of power supply to the wheels. Several results, which include consideration of varying grades and speed restrictions, were presented. However, nowadays many locomotives or motor cars can provide a continuous traction and braking force making the use of continuous-input models necessary. For a continuous-input model, Howlett et al. [15] gave necessary conditions for optimality and developed efficient numerical algorithms for calculating the optimal control inputs. Furthermore, in [16, 17] a new method is proposed to calculate the optimal switching points on a track with steep gradients. Alrecht et al. [18] used perturbation analysis to prove that the optimal switching points are uniquely defined for each steep section of track and deduced that the global optimal strategy is unique. Khmelnitsky [19] described the mathematical model of the train using the kinetic energy as the state variable. In that study, the optimal control problem was solved under varying grade profile and speed restrictions of rail lines. Liu and Golovicher [3] developed an analytical approach which combined the Pontryagin’s principle and some algebraic equations to obtain the optimal solution, which contains the sequence of optimal controls and the change points, for the continuous-input model.

The optimal trajectory of an analytical solution typically contains four optimal control regimes: maximum acceleration, cruising at constant speed, coasting, and maximum deceleration. It is worth to note that the analytical methods often meet difficulties if more realistic conditions are considered that introduce complex nonlinear terms into the model equations and the constraints [20].

- Numerical optimization
  A number of advanced techniques such as fuzzy and genetic algorithms have been proposed to calculate the optimal reference trajectory for trains. Chang and Xu [21] proposed a modified differential evolution algorithm to optimally tune the
fuzzy membership functions that provide a trade-off between punctuality, riding comfort, and energy consumption. The implementation of a genetic algorithm to optimize the coasting regions along a line is presented by Chang and Sim [22]. Han et al. [23] also used a genetic algorithm to construct the optimal reference trajectory taking non-constant grade profile, curve, and speed limits into account. They concluded that the performance of their genetic algorithm is better than that of the analytic solution obtained by Howlett and Pudney [6] in view of energy saving.

The train optimal control problem was solved by nonlinear programming and dynamic programming in [8]. The performance of a sequential quadratic programming algorithm and discrete dynamic programming were evaluated. Ko et al. [20] applied Bellman’s dynamic programming to optimize the optimal reference trajectory. Multi-parametric quadratic programming was used to calculate optimal control laws for trains in [24]. The nonlinear train model with quadratic resistance was approximated by a piecewise affine function. The resulting optimal control law was a piecewise affine function, which relates the traction force to the train position and speed.

Due to the comparable high computing power available nowadays, more and more researchers are applying numerical optimization approaches to the train optimal control problem. However, a disadvantage of numerical solution methods is that the optimal solution is not always guaranteed and the convergence speed is uncertain in general.

2.2.2 Optimal Trajectory Planning of Multiple Trains

The solution approaches for the trajectory planning of a single train presented in Sect. 2.2.1 ignore the impact caused by signaling systems, e.g., an FBS system or an MBS system. In the literature, Lu and Feng [25] considered the operation of two trains on a same line and optimized the trajectory of the following train considering the constraints caused by the leading train in an FBS system. More specifically, a parallel genetic algorithm was used to optimize the trajectories for the leading train and the following train, resulting in a lower energy consumption [25]. Gu et al. [26] also considered the trajectory planning of two trains and they applied nonlinear programming to optimize the trajectory for the following train, where two situations of the leading train, i.e., running and stopped, were considered. In addition, Ding et al. [27] took the constraints caused by an MBS system into account and developed an energy-efficient multi-train control algorithm to calculate the optimal trajectories. Three optimal control regimes, i.e., maximum traction, coasting, and maximum braking, were adopted in the algorithm and the sequences of these three regimes were determined by a predefined logic [27].

For optimal trajectory planning of trains, the analytical methods often meet difficulties to find analytical solutions if more realistic conditions are considered that introduce complex nonlinear terms into the model equations and the constraints.
The numerical optimization approaches are used more and more with the increasing computing power even though the optimal solution is not always guaranteed. In Chaps. 3 and 4 of this book, we will develop efficient approaches to provide a balanced trade-off between accuracy and computational efficiency for the trajectory planning of trains. Furthermore, since the operation of trains is highly influenced by signaling systems and only a few researchers studied the impact of signaling systems in trajectory planning problem, we will also investigate the trajectory planning problem with signaling constraints in this book.

2.3 Urban Rail Transit Scheduling Process

A general scheduling both for interurban and urban rail transit systems is a highly complex process, which is often divided into several steps [28]: demand analysis, line planning, train scheduling, rolling stock planning, and crew scheduling as shown in Fig. 2.5. First, the passenger demand has to be assessed and analyzed. Consequently, the amount of travelers wishing to go from certain origins to destinations is determined. Next, line planning is performed, which decides the routes or lines to be operated and the nominal frequency of the service. During the train scheduling step, all departure and arrival times at all stations of the lines are planned, i.e., the timetable is determined. The rolling stock planning assigns trains to all the lines. Similarly, the crews are distributed to different trains through the crew scheduling. If the crew schedule is not feasible, then the rolling stock plan or even the train schedule should be adapted. In addition, some researchers have studied the integration of steps, e.g., IVU plan introduces the integration of rolling stocking planning and crew scheduling to achieve optimal deployment of resources [29].

For urban rail transit systems, not all steps are equally important. There are specific characteristics for urban rail transit systems. The degree of freedom in the line planning is limited because the routes for the operation of trains have been fixed.

![Diagram of hierarchical planning process of railway system](image-url)
when the urban rail lines were constructed, i.e., trains do not move from one line to another during regular operation. Therefore, only the frequencies of the service, the stop-skip schedule on a certain line, and the size of train fleet can be regulated through coupling or decoupling of multiple train units to adapt varying passenger demands in urban rail transit lines. Therefore, in this section, the passenger demand and the train scheduling for urban rail transit systems will be discussed in detail.

2.3.1 Passenger Demand

Passenger demand estimation is the basis for the whole planning process. Traditionally, demand estimation relies heavily on costly and unreliable manual data collection, e.g., using passenger surveys to estimate origin-destination (OD) travel patterns. The results obtained by this kind of manual data collection maybe subject to bias and even error [30]. However, nowadays most urban rail transit systems have been equipped with automatic passenger counting systems and automatic fare collection systems, which can provide accurate passenger information to rail operators. Automatic passenger counting systems are used to count the number of passengers getting on and getting off trains at stations. With automatic fare collection systems, passengers need to use their fare cards when entering and exiting urban rail transit systems, so the location and time of each passenger’s fare transactions can be recorded.

Due to historical reasons and the complexity of optimization models, the passenger demand is usually described in one of the following two ways in the literature:

- OD-independent passenger demands
  When describing the passenger demand in an OD-independent way, the origin and destination of each passenger are not considered. The passenger arrival rate at a certain station is then e.g. defined as the number of passengers arriving at the station during a predefined time period [31].

- OD-dependent passenger demands
  The OD-dependent passenger demand is defined as an estimation of the number of people wishing to travel from an origin to a destination over a certain period of time during the day. The OD-dependent passenger demand can be conducted using the available passenger information, see [30, 32, 33] for details.

2.3.2 Train Scheduling

Train scheduling for interurban rail transit systems has been studied for decades via different mathematical techniques [34], such as linear programming [35, 36], integer or nonlinear programming [28, 37–39], and graph theory [40]. In interurban rail transit systems, the available resources, e.g., the single tracks and the crossings, are shared by trains with different origins and destinations. Thus, the trains may
overtake and cross each other at some specific locations, such as sidings and crossings. However, urban rail transit systems have the following characteristics: (1) the urban rail transit lines are usually separate from each other and have double tracks, where each track is used for one direction of the train operations, (2) train overtakings and crossings are normally not allowed during the operations, (3) the frequency of train services is much higher when compared with interurban rail transit systems. Here, we concentrate on urban rail transit systems.

2.3.2.1 Scheduling of Trains for Urban Rail Transit

In 1980, Cury et al. [41] presented a methodology to generate optimal schedules for metro lines based on a model of the train movements and of the passenger behavior. The performance index included passenger delay, passenger comfort, and the efficiency of the operation of trains. The resulting nonlinear scheduling problem was recast into several subproblems by Lagrangian relaxation and then solved in a hierarchical manner [41]. Since the convergence rate of the hierarchical decomposition algorithm can be quite poor in some cases, Assis and Milani [42] proposed a model predictive control algorithm based on linear programming to optimize the train schedule. The algorithm proposed in [42] can effectively generate train schedules for the whole day. Kwan and Chang [43] applied a heuristic-based evolutionary algorithm to solve the train scheduling problem, where the operation costs and the passenger dissatisfaction are included in the performance index. The train scheduling problem is formulated as a periodic event-scheduling problem based on a graph model in [44], which is then solved using integer programming methods. The approach proposed by Liebchen has been applied in Berlin subway systems [45]. The passenger transfer behavior and transfer waiting times are considered in [46], which presents a mixed-integer programming optimization model to synchronize the train schedules for different urban rail transit lines. Furthermore, a demand-oriented timetable design is proposed in [47], where the optimal train frequency and the capacity of trains are first determined and then the schedule of trains are optimized. Vazquez et al. [48] proposed a stochastic approximation approach to adjust the frequencies of different urban transit lines according to the observed variable passenger demand. However, the energy consumption of railway operation and dwell times at stations are not included in the model of [48].

2.3.2.2 Real-Time Scheduling or Rescheduling of Trains

Since trains do not run exactly according to the predefined schedule in practice, real-time scheduling approaches have been proposed. In the literature, there are several interpretations for real-time scheduling. For interurban railway systems, real-time scheduling is based on the existing timetable data and is used to handle route conflicts due to train delays or incidents [40, 49–56]. However, in urban rail transit systems,
real-time scheduling regulates the headways between trains based on a train schedule with a constant headway.

Several rescheduling approaches have been proposed for urban rail transit systems [57]: holding, zone scheduling, short turning, deadheading, and/or stop-skipping [31, 58–60]. Holding is used to regulate the headways by holding an early-arriving train, or a train with a relatively short leading headway [31]. In zone scheduling [58], the whole line is divided into several zones, where the trains stop at all stations within a single zone and then run to the terminal station without stopping. The required number of trains and drivers and passenger travel times may be reduced by the zone scheduling, where the zones are defined based on the passenger flows. There are short-turning and full-length trips operating on the line in the short-turning strategy [59, 60], where the short-turning trips serve only the zone with high demands and the full-length trips run the whole line. The deadheading strategy involves some trains running empty through a number of stations at the beginning of their trips to reduce the headways at later stations [57, 61]. A dynamic stop-skipping strategy is frequently used in lines with high demands, as it allows those trains that are late and behind the schedule to skip certain low-demand stations and in that way increase the running speed.

Wong and Ho [62] proposed dwell time and running time control for the real-time rescheduling problem of urban rail transit systems. They applied a dynamic programming approach to their rescheduling model to devise an optimal set of dwell times and running times [62]. In addition, Goodman and Murata [63] formulated the train rescheduling problem from the perspective of passengers, where a gradient calculation method was developed to solve the rescheduling problem in real time. Furthermore, Norio et al. [64] proposed to use passenger dissatisfaction as a criterion for the rescheduling and applied a meta-heuristics algorithm to solve the rescheduling problem.

As demonstrated in [65, 66], the stop-skipping strategy can reduce the passenger travel time and the operation cost of rail transit operators. The stop-skipping operation was first developed for the Chicago metro system in 1947 [65]. Now, the SEPTA line in Philadelphia, Helsinki commuter rail, and the metro system in Santiago, Chile apply the stop-skipping train schedule in practice. They apply a static stop-skipping strategy [66], i.e., the A/B skip-stop strategy, where stations are divided into three types: A, B, and AB; A train services stop at A stations and AB stations, while B train services stop at B stations and AB stations. Major stations are usually labeled with the type AB; so all trains stop there. The transit operators provide the stop-skipping information to passengers via panels at platforms and announcements in the trains. The Santiago metro operator stated that passengers adapt to the stop-skipping strategy quickly [65]. Elberlein [67] formulated the stop-skipping problem as a mixed integer nonlinear programming problem (MINLP), where trains can skip some station strings (i.e., a collection of consecutive stations). Fu et al. [57] represented the skipping of stations by trains as binary variables and obtained a MINLP problem, which was solved using an exhaustive approach. Lee [66] applied genetic algorithm to obtain the optimal train schedule and to find the best combination of the stop-skipping trains and the all-stop trains based on the A/B stop-skipping strategy.
The passenger demand for urban rail transit systems increases dramatically and varies significantly along urban rail transit lines and the time of the day. To satisfy the passenger demand, trains are operated with small headway, which is around 2–5 min. Therefore, the scheduling of trains according to the passenger demand becomes more and more important for reducing the operation costs and for guaranteeing passenger satisfaction. In particular, the passenger satisfaction can be characterized by waiting times at platforms, onboard travel times, the number of transfers, the onboard crowdedness, etc.

2.4 Summary

A brief introduction to the operation of trains and the principle of signaling systems has been presented in this chapter. We have briefly discussed the literature of the optimal trajectory planning for trains and of the train scheduling for urban rail transit systems. In addition, we have motivated why the work of this book is needed.

References

29. IVU.rail (2015) Integrated scheduling, dispatching and optimization
20 2 Background: Train Operations and Scheduling


Optimal Trajectory Planning and Train Scheduling for Urban Rail Transit Systems
Wang, Y.; Ning, B.; van den Boom, T.; De Schutter, B.
2016, XXI, 180 p. 59 illus., 2 illus. in color., Hardcover
ISBN: 978-3-319-30888-3