A survey on optimal trajectory planning for train operations

Y. Wang, B. Ning, F. Cao, B. De Schutter, and T.J.J. van den Boom

If you want to cite this report, please use the following reference instead:

Abstract—Because of the rising energy prices and environmental concerns, the calculation of energy-optimal reference trajectories for trains is significant for energy saving. On the other hand, with the development automatic train operation (ATO), the optimal trajectory planning is significant to the performance of train operation. In this paper, we present an integrated survey of this field. First, a nonlinear continuous-time train model and a continuous-space model of train operations are described, after which the optimal trajectory planning problem is formulated based on these two models. The various approaches in the literature to calculate the reference trajectory are reviewed and categorized into two groups: analytical solutions and numerical optimization. Finally, a short discussion of some open topics in the field of optimal trajectory planning for train operations are given.

Index Terms—train operations, optimal trajectory planning, energy efficient

I. INTRODUCTION

Because of the rising energy prices and environmental concerns, the energy efficiency of transportation systems becomes more and more important [1]. In the US, about 28% of the country energy consumption is represented by the transportation [2]. Similarly, the energy consumed by transportation is approximately 30% of the UK energy consumption [3]. Furthermore, it is well known that railway transport combines high transport capacity and high efficiency. It plays a more and more important role for public transportation in the near future. The reduction of energy consumption is one of the key objectives of railway systems. Meanwhile, the interest of railway operators in energy efficiency has been rising more and more in recent years. Even a small improvement in energy consumption can make the railway operators save a lot of money. Therefore, some systems have been developed to assist drivers to drive the train optimally, such as FreightMiser [4], Metromiser [4], and driving style manager [5].

With the development of modern railway systems, automatic train control system has become vital equipment that ensures the running safety, shortens the train headway, and improves the quality of train operation [6]. An automatic train control system consists of an automatic train protection system, an automatic train supervision system, and an automatic train operation (ATO) system. ATO system plays a key role in ensuring accurate stopping, operation punctuality, energy saving, and riding comfort [6]. It is an important unit of the automatic train control system responsible for controlling the train speed to achieve safe and reliable operation. A typical ATO system consists of two levels of control actions, as conceptually illustrated in Figure 1. The high-level control involves the optimal control problem that is responsible for calculating the optimal speed-position reference trajectory based on the information collected by automatic train control systems, such as line resistance, speed limits, maximum traction and braking forces, etc. The low-level control is used to make the train track the pre-planned reference trajectory via certain control methods.

The driving performance including punctuality, energy consumption, etc. strongly depends on the optimal reference trajectory both when the train is under driver control or controlled by the ATO system. The FreightMiser and Metromiser systems mentioned before were developed by the scheduling and control group (SCG) 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 [4]. The SCG researchers mainly focused on minimizing the energy consumption by Pontryagin’s principle which is also known as maximum principle [4]. The driving style manager developed by Bombardier implements
discrete dynamic programming to obtain optimal results, which are then displayed to the train driver [5]. When the train stops at a station, the driving style manager calculates the optimal trajectory to the subsequent station using real-time information. For an ATO system, the calculation of the optimal trajectory is the core of the high-level control. Therefore, a feasible and efficient algorithm for calculating the reference trajectory is significant to the driving performance.

The research on optimal trajectory planning of train operations began in the 1960s. A simplified train optimal control problem was studied by Ichikawa [7], who solved the problem by using Pontryagin’s principle. Later on, a lot of 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 the following two main categories [5]:

- Analytical solution.
- Numerical optimization.

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

The rest of this paper is structured as follows. In Section II the nonlinear continuous-time model and continuous-space model for trajectory planning are presented. Furthermore, the optimal trajectory planning problems based on these two models are formulated. In Section III the research based on an analytical approach is reviewed. The numerical optimization approaches used for optimal trajectory planning are investigated in Section IV. We conclude with a short discussion of some issues for future work in Section V.

II. PROBLEM DEFINITION

A. Train model

In the literature of train optimal trajectory planning, there are two approaches to describe train operations: the single mass-point approach and the distributed mass approach. The distributed mass approach was proposed by Howlett and Pudney [4]. In addition, they showed that the optimal trajectory planning problem by using a distributed mass approach can be equivalent to the one by using the single mass-point approach. However, the single mass-point approach for a train is often used in the literature [8]. In this setting, the motion of a train can be described by the following simple continuous-time model [2]:

\[
m \rho \frac{dv}{dt} = u(t) - R_b(v) - R_\ell(s,v)
\]

(1)

\[
\frac{ds}{dt} = v
\]

(2)

where \( m \) is the mass of the train, \( \rho \) is a factor to consider the rotating mass [1], \( v \) is the velocity of the train, \( s \) is the position of the train, \( u \) is the control variable, i.e. the traction or braking force, which is bounded by the maximum traction force \( u_{\text{max}} \) and the maximum braking force \( u_{\text{min}} \). \( R_b(v) \) is the basic resistance including roll resistance and air resistance, \( R_\ell(s,v) \) is the line resistance caused by track grade, curves and tunnels.

The maximum traction force \( u_{\text{max}} \) is often considered as a constant in the literature [9]. However, in reality it is a function of velocity \( v \). Due to the maximum adhesion and the characteristics of the power equipment [1], the diagram of the maximum traction force \( u_{\text{max}}(v) \) normally looks like the one shown in Figure 2 [10]. This diagram is described as a group of hyperbolic or parabolic formulas in [1], where each formula approximates the actual traction force for a certain speed interval. According to the arguments for the maximum braking force given in [1], the full braking effort is reserved for an emergency stop. Under normal circumstances the train driver or automatic train operation system brakes in a comfort mode, where the maximum force for the service breaking is 0.75 times that of the emergency braking, i.e. the full braking effort. On the other hand, the braking effort (including the maximum braking effort) is considered as a constant by some common safety systems, such as the European Train Control System and the German continuous train control system [1]. Therefore, the maximum force for service braking is usually considered as a constant.

In practice, according to the Strahl formula [11] the basic resistance \( R_b(v) \) can be described as

\[
R_b(v) = m(a_1 + a_2v + a_3v^2)
\]

(3)

where \( m \) is the train mass, the coefficients \( a_1, a_2 \), and \( a_3 \) depend on the train characteristics and the wind speed, which can be calculated from the data known about the train.

The line resistance \( R_\ell(s,v) \) caused by track slope, curves, and tunnels can be described by [12]

\[
R_\ell(s,v) = mg \sin(\alpha(s) + f_c(r(s)) + f_l(l(s),v)
\]

(4)

where \( g \) is the gravitational acceleration, \( \alpha(s) \), \( r(s) \) and \( l(s) \) are the slope, the radius of the curve, and the length of the tunnel along the track, respectively. The curve resistance \( f_c(\cdot) \) and the tunnel resistance \( f_l(\cdot) \) are given by empirical formulas. An example empirical formulas of the curve resistance is Roeckl’s formula [13]

\[
f_c(r(s)) = \frac{6.3}{r(s) - 55}m \quad \text{for} \quad r(s) \geq 300m
\]

(5)
When running in tunnels, the train experiences a higher air resistance that depends on the tunnel form, the smoothness of tunnel walls, the exterior surface of the train, and so on. The expression of tunnel resistance is [13]
\[ f_t(l(s), v) = a_t(l(s))v^2, \]  
where \( a_t \) is the tunnel factor, which depends on tunnel length and train type. For the tracks outside the tunnels, the coefficient \( a_t \) is equal to zero. For more complete information on the calculation of \( a_t \), see the paper written by Vardy and Reinke [14].

### B. Optimal Control Problem based on the time domain

As stated in [2], reference trajectory planning for trains can be formulated as one of the problems in optimal control theory. The traction or braking force \( u \) then is the control variable, which is calculated to satisfy conditions and restrictions on the control variable and on the state variables, i.e. train position \( s \) and speed \( v \). The objective function could be the minimum time, or the energy consumption for a given trip time, or the total operation cost, such as a weighted sum of electrical energy consumption and trip time. In addition, the riding comfort can also be considered, which is expressed as a function of the change of the control variable \( u \) since reducing the number of transitions and the rate of change of \( u \) may improve passenger comfort [15]. The objective function can then be written as:

\[ J = \int_0^T u(t) \cdot v(t) + \lambda \cdot \left| \frac{du(t)}{dt} \right| dt \rightarrow \min \]  
subject to train dynamics (1), the following constraints
\[ u_{\text{min}} \leq u(t) \leq u_{\text{max}} \]  
\[ 0 \leq v(t) \leq V_{\text{max}}(s) \]  
and the boundary conditions
\[ s(0) = s_{\text{start}}, \quad v(0) = v_{\text{start}}, \]  
\[ s(T) = s_{\text{end}}, \quad v(T) = v_{\text{end}}, \]  
where \( J \) is the weighted sum of the energy consumption and riding comfort of the train operation, \( \lambda \) is the weight, the maximum allowable velocity \( V_{\text{max}}(s) \) depends on the train characteristics and the line conditions (and as such it is usually a piecewise constant function of the coordinate \( s \) [2], [16]). \( s_{\text{start}} \) and \( v_{\text{start}} \) are the coordinate and the velocity at the beginning of the route, and \( s_{\text{end}} \) and \( v_{\text{end}} \) are the coordinate and the velocity at the end of the route. The duration of the trip \( T \) is usually given by the timetable.

### C. Optimal control problem based on the position domain

As proposed in some previous works [2], [8], [16], [9], some researchers stated that it is better to choose the coordinate \( s \) as an independent variable rather than time \( t \). On the one hand, the choice of position \( s \) as independent variable will simplify the consideration of track-related data, such as line resistance and speed limits. On the other hand, the analytical and numerical study of the optimal problem will be significantly simplified. Furthermore, Khmelitsky [16] chose the total energy of the train and the time \( t \) as states where the total energy includes kinetic and potential energy. Similarly, Franke et al. [8] used the kinetic energy per mass unit and the time as states. The choice of kinetic energy instead of the speed \( v \) facilitates the study of the optimal control problem, because this choice eliminates some of the model nonlinearities. So in [8] the kinetic energy per mass unit \( E = 0.5v^2 \) and the time \( t \) are chosen as states, and the coordinate \( s \) as the independent variable. The continuous-time model (1) and (2) can then be rewritten as the following continuous-space model:

\[ mp^2 \frac{d^2 s}{ds^2} = u(s) - R_0(\sqrt{2E}) - R_1(s) \]  
\[ \frac{dt}{ds} = \frac{1}{\sqrt{2E}} \]  

The optimal control problem (8)-(12) can be stated as:

\[ J = \int_{s_{\text{start}}}^{s_{\text{end}}} u(s) + \lambda \left| \frac{ds}{ds} \right| ds \rightarrow \min \]  
subject to the model (13) and (14), the following constraints
\[ u_{\text{min}} \leq u(s) \leq u_{\text{max}} \]  
\[ 0 \leq E(s) \leq E_{\text{max}}(s) \]  
and the boundary conditions, which are rewritten as
\[ E(s_{\text{start}}) = E_{\text{start}}, \quad E(s_{\text{end}}) = E_{\text{end}} \]  
\[ t(s_{\text{start}}) = 0, \quad t(s_{\text{end}}) = T \]  
where \( E_{\text{max}}(s) = 0.5v_{\text{max}}^2(s) \), \( E_{\text{start}} = 0.5v_{\text{start}}^2 \), and \( E_{\text{end}} = 0.5v_{\text{end}}^2 \). An assumption should be noted for the above equations: it is assumed that the \( E(s) > 0 \), which means the train’s speed is always strictly larger than zero, i.e. the train travels nonstop. Khmelnitsky [16] states that this assumption is not restrictive in practice for two reasons. First, the speed of the initial start and the terminal stop can be approximated by small nonzero velocities. Second, stops at an intermediate point of the trip will not be planned deliberately in the optimal control design for a single train’s operation.

### III. Analytical solution

According to whether the traction and braking force is continuous or discrete, there are two kinds of solution approaches. One kind of approaches is for train operations with a continuous input, while the other one is for train operations with a discrete input [9].
A. The optimal trajectory planning with a discrete input

When train operations with a discrete input, Only a few finite number of pre-determined values can be taken by the control variable $u$. Nowadays, only some freight diesel locomotive is still with discrete traction and braking force [2], whose throttle can take only a finite number of positions and each position of the throttle determines a constant rate of fuel supply [4]. In each position, the power produced by the locomotive is directly proportional to the rate of fuel supply. The research on trajectory planning for a train with a discrete input is mainly done by the SCG of the University of South Australia.

The research on optimal trajectory planning for a train with a discrete input was inspired by the early research of Howlett [17] for a train with continuous input. Howlett [17] simplified the analysis by using Pontryagin’s principle to reformulate the problem as a finite dimensional constrained optimization where the unknown switching times are the variables. For a train with a discrete input, the Karush-Kuhn-Tucker conditions are used to derive the optimal trajectory for each fixed control sequence by finding the switching times [9]. Howlett and Leizarowitz [18] obtained an algebraic equation, rather than a differential equation, from the Euler-Lagrange equation for certain intervals. This algebraic equation is useful for the structure of optimal control scenarios which are composed of segments with pure control and segments with chattering control.

The optimal trajectory planning problem with a discrete input was studied for a track without varying gradient and speed limit by Cheng and Howlett [19]. They showed that an optimal trajectory depended on three critical values of the velocity for a prescribed sequence of fuel supply rates. Then, the problem with speed limits was solved by Pudney and Howlett [20]. They obtained that on intervals of track where the speed limit is below the desired cruising speed, the speed must be held at the limit. The later on the problem with continuously varying gradient was solved by Howlett and Cheng [21]. Afterwards, the problem with non-zero track gradient and speed limit was solved by Cheng [22], which was difficult to find analytic solution because it is no longer possible to follow an arbitrary smooth speed limit precisely [9].

B. The optimal trajectory planning with continuous-input

Nowadays a lot of locomotives can provide a continuous traction and braking force. For a continuous-input train, Khmelnitsky [16] gave a comprehensive analysis of the optimal trajectory planning with a continually varying gradient and speed restrictions. As stated before, he described the mathematical model of the train by using the total energy and the time as the state variables. The state equations are essentially the same as (13) and (14). The Pontryagin’s principle was applied to obtain the analytical properties about optimal control scenarios and their sequences. Khmelnitsky proved that under certain conditions there exists a unique optimal solution. The calculation complexity depends on the complexity of the grade profile and speed restrictions. Liu and Golovitcher [2] stated that the approach proposed by Khmelnitsky is less effective than the approach of Golovitcher [23], because the procedure of Khmelnitsky includes integration of an additional differential equation for the conjugate function. The calculation of the reference trajectory in [23] is also based on the Pontryagin’s principle, but the general rules for building the sequence of optimal control scenarios and necessary optimality conditions are simpler algebraic equations. So this approach is very efficient and the mathematical model is the same. Based on the former research, a complete solution of the optimal trajectory planning was given by Liu and Golovitcher [2].

Howlett [9] applied the Pontryagin’s principle to find necessary conditions for the reference trajectory. He determined the optimal switching points by using the key equations yielded by these necessary conditions. In addition, Howlett and Pudney [4] showed that any reference trajectory of the continuous input system can be approximated as closely as possible by a reference trajectory of the discrete input system. Recently, new formulae were developed by Vu [24], which were used to continually update reference trajectories on-line. Vu stated that the global optimal control strategy with the critical switching points can be obtained by applying a new local energy minimization principle over each steep section separately.

Based on analytical approaches mentioned above, there are four optimal control scenarios in the optimal trajectory: maximum acceleration, cruising at constant speed, coasting, and maximum deceleration. However, a more detailed model of train operation is considered by Franke et al. [5], which includes the efficiency of the propulsion system and regenerative brake scheme of electrical rail vehicle. The efficiency of the propulsion system is usually considered as a constant in literature. In reality, it actually varies greatly with the operating conditions. The power losses of the propulsion system is a nonlinear function of the traction force $u$ and the velocity $v$ (see Figure 1 in [5]). As stated in [5], with the consideration of the efficiency of the propulsion system, no maximum acceleration or maximum deceleration is applied at high velocities, because when maximum traction and braking force is applied at the high speeds, the power loss is large. So there is a trade-off between the speed $v$ and the control variable $u$.

It is worth to note that the analytical methods often meet difficulties to find the analytic solution if more realistic conditions are considered that introduce complex nonlinear terms into the model equations and the constraints [25].

IV. NUMERICAL OPTIMIZATION

Liu and Golovitcher [2] stated that classical numerical optimization methods were not feasible to solve the optimal trajectory planning problem on an on-board computer for real-time calculations, because these methods, such as discrete dynamic programming, required significant calculation time. Therefore, the research in this field was impeded for a long time because of the computation difficulties. However, due to the high computing power available nowadays, more
and more researchers are applying numerical optimization approaches to obtain the reference trajectory.

A. Fuzzy and evolutionary algorithms

A number of advanced techniques such as fuzzy and genetic algorithms have been proposed to calculate the optimal reference trajectory for train operation. Yasunobu et al. [26] proposed a fuzzy ATO controller and implemented it in Sendai of Japan in 1987. This controller can control each train’s departure, speed regulation, and dwell time. Membership function plays an important role in ensuring the control accuracy and robustness of the fuzzy ATO controller. So Chang and Xu [15] proposed a modified differential evolution algorithm to optimally tune the fuzzy membership functions that provide a trade-off among punctuality, riding comfort, and energy consumption.

The implementation of a genetic algorithm to optimize the coast control was demonstrated by Chang and Sim [27]. The optimal results in [27] are given as a coast control tables which is referenced by ATO system for deciding when to initiate coasting and resume maximum acceleration. Han et al. [28] also use a genetic algorithm to construct the optimal reference trajectory. They defined the switch points of the optimal control scenarios as the chromosome string and the energy consumption as the fitness function. Han et al. [28] conclude that the performance of their genetic algorithm is better than the analytic solution obtained by Howlett and Pudney [4] in view of energy cost. A formal method combined genetic algorithm and fuzzy logic was proposed in [29] to optimize a weighted sum between energy consumption and running time. Bochannikov et al. [29] concluded that the energy saving was affected by the acceleration and deceleration rates by running a series of simulations in parallel by using genetic algorithm. Acikbas and Soylemez [30] combined artificial neural networks and genetic algorithm to obtain the optimal coasting speed. The objective function is considered as a weighted sum of energy consumption and the cost of running time.

B. Dynamic programming

Some decision variables of the analytical solution can be obtained by using gradient search, such as cruising speeds, switching points [5]. The combination of the analytical solution and gradient search has been successfully applied by some research [5].

Nowadays, computation power has increased considerably compared to the period which most of the papers mentioned in section III were written. Therefore, Franke et al. [5] proposed a more detailed nonlinear train model, in which the power loss of the inverter locomotive is modeled. The optimal trajectory planning problem based on this nonlinear model is solved by nonlinear programming and dynamic programming [5]. It is concluded that discrete dynamic programming turned out to be better to deal with the nonlinear optimal problem compared to sequential quadratic programming, because the overall calculation time of discrete dynamic programming is deterministic and the result of the computation is obtained in the form of a feedback control law. Franke et al. [5] states that no maximum acceleration or maximum deceleration is applied at high velocities when considering the power loss of the propulsion system (see also the end of section III-B).

Ko et al. [25] apply Bellman’s dynamic programming to optimize the optimal reference trajectory. The original problem is then transformed into a multi-state decision process. Dynamic programming, a gradient method, and sequential quadratic programming are introduced to solve the optimal trajectory planning problem in [31]. Under simple and complicated operation conditions, the simulations showed the gradient method had good convergence.

Multi-parametric quadratic programming is used in [32] to calculate the optimal control law of train operation. The nonlinear train model with quadratic resistance is approximated by a piecewise affine function. The resulting optimal control law is a time-varying piecewise affine function, which relates the traction force to the train position and speed. Therefore, this is a off-line computed optimal feedback control policy that can be easily evaluated on-line.

However, the optimal solution is not always guaranteed in these numeric optimization approaches. Since, the “optimal” solution obtained could be a local minimum. In addition, the convergence speed is uncertain in general. Moreover, the computation of these numeric optimization approaches is often too slow for real-time application, e.g. the computation time in [32] is 12 hours.

V. DISCUSSION

The optimal trajectory planning problems based on the nonlinear continuous-time model and continuous-space model of train operations are presented in this paper. The various methods in the literature are grouped into these two main categories: analytical solution and numerical optimization.

As stated above, 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. For the numerical optimization approaches, the optimal solution is not always guaranteed. In addition, the computation is often too slow. So the following open topics for future work arise.

First, there are some efficient optimal control software packages using nonlinear programming available so there is a trade-off involved between optimality and computational speed, such as DIDO [33], GPOPS [34], PROPT [35]. New approaches can be explored to solve this optimal trajectory planning problem by using these efficient optimal control softwares. Moreover, it would interesting and highly relevant

\[ \min_x x^T H x + (C + q^T E)x \]

s.t. \[ Ax \leq b + Dq \]

with \( q \) a parameter.
to obtain guarantees on the convergence speed and on the degree of sub-optimality of the solution found. In the latter content, one option is to solve the optimal trajectory planning problem as a mixed integer linear programming (MILP) problem, which can be solved efficiently using existing commercial and free solvers [36], [37]. The optimal solution of the MILP problem is the global minimum and always guaranteed. However, we have to make some approximations to construct an MILP formulation of the nonlinear train operation model.

In our view, significant progress can be achieved in the optimal trajectory planning domain for train with the development of efficient approaches that provide a good trade-off between accuracy and computational efficiency.

REFERENCES
