Pseudospectral optimal train control

Goverde, Rob M.P.; Scheepmaker, Gerben M.; Wang, Pengling

DOI
10.1016/j.ejor.2020.10.018

Publication date
2021

Document Version
Final published version

Published in
European Journal of Operational Research

European Journal of Operational Research 292 (2021) 353–375

Contents lists available at ScienceDirect

European Journal of Operational Research

journal homepage: www.elsevier.com/locate/ejor

Interfaces with Other Disciplines

Pseudospectral optimal train control

Rob M. P. Goverde[a,*], Gerben M. Scheepmaker[a,b], Pengling Wang[a]

[a] Department of Transport and Planning, Delft University of Technology, Delft, the Netherlands
[b] Netherlands Railways, Department of Performance management and Innovation, Utrecht, the Netherlands

ARTICLE INFO

Article history:
Received 8 August 2019
Accepted 13 October 2020
Available online 30 October 2020

Keywords:
Optimal train control
Train trajectory optimization
Pontryagin's Maximum Principle
Pseudospectral method
Singular solution

ABSTRACT

In the last decade, pseudospectral methods have become popular for solving optimal control problems. Pseudospectral methods do not need prior knowledge about the optimal control structure and are thus very flexible for problems with complex path constraints, which are common in optimal train control, or train trajectory optimization. Practical optimal train control problems are nonsmooth with discontinuities in the dynamic equations and path constraints corresponding to gradients and speed limits varying along the track. Moreover, optimal train control problems typically include singular solutions with a vanishing Hessian of the associated Hamiltonian. These characteristics make these problems hard to solve and also lead to convergence issues in pseudospectral methods. We propose a computational framework that connects pseudospectral methods with Pontryagin's Maximum Principle allowing flexible computations, verification and validation of the numerical approximations, and improvements of the continuous solution accuracy. We apply the framework to two basic problems in optimal train control: minimum-time train control and energy-efficient train control, and consider cases with short-distance regional trains and long-distance intercity trains for various scenarios including varying gradients, speed limits, and scheduled running time supplements. The framework confirms the flexibility of the pseudospectral method with regards to state, control and mixed algebraic inequality path constraints, and is able to identify conditions that lead to inconsistencies between the necessary optimality conditions and the numerical approximations of the states, costates, and controls. A new approach is proposed to correct the discrete approximations by incorporating implicit equations from the optimality conditions. In particular, the issue of oscillations in the singular solution for energy-efficient driving as computed by the pseudospectral method has been solved.

1. Introduction

Optimal control theory is widely applied in different fields to find the controls that minimize a cost functional subject to dynamic constraints, path constraints and boundary conditions. One of the applications of optimal control theory is *minimum-time train control* (*MTTC*) with the aim to compute the minimum time between two train stops. Another important application is *energy-efficient train control* (*EETC*) or also referred to as *energy-efficient train trajectory optimization*, with the aim to minimize total traction energy. The train controls consist of traction and braking. These optimal train control problems are highly complex in practice with discontinuous dynamic equations, pure state and control algebraic inequality constraints, and mixed state-control algebraic inequality constraints. Moreover, the solution to the EETC problem generally contains singular arcs. Hence, solving optimal train control problems is quite challenging, and specifically EETC problems. Most research on optimal train control has focused on the application of *Pontryagin's Maximum Principle* (*PMP*) (Pontryagin, Boltyanskii, Gamkrelidze, & Mishchenko, 1962). The application of this theory leads to the optimal driving regimes consisting of maximum acceleration, cruising, coasting and maximum braking. The challenge is to determine the optimal switching structure with the exact switching points as well as the optimal sequence of the driving regimes. Scheepmaker, Goverde, and Kroon (2017) provided a recent extensive literature review on EETC. One of the conclusions was that the pseudospectral method was a promising approach to solve generic EETC problems with varying track gradients and speed limits. In this paper, we further explore the Radau pseudospectral method to optimal train control based on structured numerical experiments that are assessed on consistency with the optimality conditions obtained by the PMP. We identify numerical

* Corresponding author.
E-mail addresses: r.m.p.goverde@tudelft.nl (R.M.P. Goverde), g.m.scheepmaker@tudelft.nl (G.M. Scheepmaker), p.l.wang@tudelft.nl (P. Wang).

https://doi.org/10.1016/j.ejor.2020.10.018

inaccuracies due to convergence issues, discretization and oscillating singular control, and propose a new approach based on combining underdetermined implicit equations from the PMP with the pseudospectral approximate solutions to get accurate continuous optimal control solutions.

Optimal control problems can be solved by indirect and direct methods (Betts, 2010; Rao, 2014). An indirect method derives the necessary optimality conditions based on PMP, which leads to a two-point boundary value problem (BVP) in the states and adjoint costates that needs to be solved by numerical methods (Ross, 2005). These methods thus indirectly solve the original optimal control problem by solving an associated BVP. In general, however, the BVP is hard to solve, as it is very sensitive to the initial guesses for the unknown (costate) boundary conditions and a priori knowledge of the switching structure of the inequality path constraints is required. Therefore, most indirect methods use (heuristic) constructive methods to find (sub)optimal driving strategies based on the PMP optimality conditions, with (linearity) simplifications or assumptions on driving regimes (e.g. no coasting, no cruising below the speed limit). Most papers on optimal train control used indirect methods based on PMP optimality conditions (Albrecht, Howlett, Pudney, Vu, & Zhou, 2016a; 2016b; Howlett & Pudney, 1995; Khmelnitsky, 2000; Liu & Golovitcher, 2003). For the energy-efficient train control problem, algebraic formulae for the costate along sections with constant gradient can be derived, which led to efficient real-time algorithms (Albrecht et al., 2016a; 2016b; Howlett, Pudney, & Vu, 2009; Liu & Golovitcher, 2003). In addition, many heuristic methods have been applied using implicit knowledge of the optimal control structure, for example Chevrier, Pellegrini, and Rodriguez (2013), Sicre, Cucala, and Fernández-Cardador (2014) and Haahr, Pisinger, and Sabbaghian (2017). For more details and references, see Scheepmaker et al. (2017).

On the other hand, direct solution methods transcribe the infinite-dimensional optimal control problem into a finite-dimensional nonlinear programming (NLP) problem using collocation of the differential equations and the integral objective functional. The resulting NLP problem is then solved using efficient nonlinear optimization algorithms (Betts, 2010). This direct approach can be used for highly complex problems, since there is no need to derive the first-order necessary optimality conditions, they are less sensitive to initial solutions, and no a priory knowledge is needed of the active and inactive inequality path constraints. In the optimal train control literature direct methods have been used only recently. Most approaches are based on pseudospectral methods (Scheepmaker & Goverde, 2016; Wang & Goverde, 2016a; 2016b; 2017; 2019; Wang, De secondchutter, Van den Boom, & Ning, 2013; Ye & Liu, 2016; 2017). Wang et al. (2013); Wang, DeSchutter, Van den Boom, and Ning (2014) also applied a mixed-integer linear programming (MILP) approach using piecewise affine functions. At this stage the computation times for the direct method are not competitive with the computation times for the indirect methods that are currently used for on-board calculations.

Over the past two decades, pseudospectral methods have become popular for solving optimal control problems within particular the aerospace domain (Garg et al., 2010; Ross & Karpenko, 2012). A pseudospectral method discretizes the state and control and then transcribes the continuous-time optimal control problem to a finite-dimensional NLP that is solved using established NLP solvers. In a pseudospectral method the state and control are approximated by global polynomials at collocation points using a basis of Lagrange (or Chebyshev) polynomials. Typical collocation points are Legendre-Gauss (LG), Legendre-Gauss-Lobatto (LGL), and Legendre-Gauss-Radau (LGR) points (Garg et al., 2010; Ross & Karpenko, 2012). These points are obtained from the roots of a Legendre polynomial and/or its derivatives. They are all defined on the domain $[-1, 1]$ but differ in that the LG points include neither of the endpoints, the LGR points include one endpoint, and the LGL points include both endpoints. Different pseudospectral methods have been developed based on the collocation points: the Gauss (LG) pseudospectral methods (Benson, Huntington, Thorvaldsen, & Rao, 2006; Rao et al., 2010), the Legendre (LGL) pseudospectral methods (Elnagar, Kazemi, & Razzaghi, 1995; Fahroo & Ross, 2001; Gong, Fahroo, & Ross, 2008a; Ross & Fahroo, 2004), and the Radau (LGR) speudospectral methods (Garg et al., 2011). The differences between the methods lie in the the degree of the Lagrange polynomial, the boundary condition and the typical problem horizon (Garg et al., 2009; Ross & Karpenko, 2012). The use of global polynomials together with Gaussian quadrature collocation points provides accurate approximations that converge exponentially fast for smooth problems (Garg et al., 2010). For optimal control problems with discontinuities in the constraints the problem can be partitioned into multiple phases, where the phase boundaries (also called mesh points or knots) can be chosen at the points of discontinuities (Betts, 2010). In these multiple-phase optimal control problems, each phase is solved with a separate set of collocation points, while additional linking conditions glue the variables of the adjacent phases together (Darby, Garg, & Rao, 2011a; Darby, Hager, & Rao, 2011b; Patterson, Hager, & Rao, 2015; Patterson & Rao, 2014; Rao et al., 2010; Ross & Fahroo, 2004). Hence, with multiple phases the state is approximated by multiple polynomials that are linked at the phase boundaries.

The pseudospectral methods can also compute estimations of the continuous costates based on the associated Lagrange multipliers computed for the discretized NLP problem (Darby et al., 2011a; Fahroo & Ross, 2001; Garg et al., 2011; 2010). For the Gauss and Radau pseudospectral methods the discrete approximations are obtained directly by a transformation of the Lagrange mutipliers using the quadrature weights of the collocation points. For the Legendre pseudospectral methods the discrete costates are not uniquely determined and need a closure condition related to the transversality conditions of the necessary optimality conditions for the continuous optimal control problem (Fahroo & Ross, 2001; Garg et al., 2010; Gong, Ross, Kang, & Fahroo, 2008b). The covector mapping principle for the Legendre Pseudospectral method then states that these multipliers of the discretized NLP problem converge to the costates of the discretized necessary optimality conditions (Gong et al., 2008a; Gong et al., 2008b; Ross & Karpenko, 2012). Comparative studies showed that the Radau pseudospectral method provides the most accurate results, including the costates, with a fast convergence rate (Garg, 2011; Garg et al., 2009; Huntington, 2007). Moreover, the LGR collocation points are most suitable for multiple-phase optimal control problems since they include one endpoint per phase, and thus all phase boundaries are collocation points corresponding to discontinuities in the input data. Together, the LGR collocation points thus cover all phases and their boundaries except the terminal point which is however estimated well by the state approximations. The Radau pseudospectral method is implemented in the MATLAB toolbox GPOPS (General Pseudospectral OPtimal Control Software) (Rao et al., 2010), which we used in this paper for the experiments.

The pseudospectral method works well for smooth problems, but the optimal train control problems may be nonsmooth in practice due to infrastructural constraints. In particular, the varying slopes of the railway tracks are usually modelled as piecewise constant gradients resulting in a discontinuous dynamic equation, and varying speed limits along the track result in discontinuous state inequality path constraints that may lead to discontinuous costate functions (Bryson & Ho, 1975). Moreover, the optimal train control problems have a Hamiltonian that is linear in the control and thus has a zero Hessian. This may cause singular arcs for which no explicit analytical expressions can be derived from the necessary optimality conditions and for which the pseudospectral so-

lutions show oscillating control and state behaviour. Furthermore, stop distances may be very long with possibly many fluctuations in gradients and speed limits leading to many phases of the resulting multiple-phase optimal control problem with as many interior state equality constraints. The literature on optimal train control using pseudospectral methods indeed shows irregularities in the control and/or state profile plots that need to be understood before these methods can be used in practice (Wang & Goverde, 2016a; 2019; Ye & Liu, 2016; 2017). Ye and Liu (2017) compared the pseudospectral method with a heuristic solution method consisting of an NLP problem formulation based on closed-form expressions for each driving regime obtained with some simplifying assumptions. For one case the pseudospectral method found the same energy consumption, despite some fluctuations in the singular cruising regime. In a second case study the heuristic method found a solution with less energy consumption and they concluded that it is quite difficult to choose appropriate parameters of GPOPS to ensure convergence within acceptable tolerances. Wang et al. (2013, 2014) compared a pseudospectral method with an MILP method. The pseudospectral method provided solutions with the least energy but was much slower than the MILP method. However, the MILP model showed a rough approximation while the pseudospectral method gave smooth results, which raises the question how accurate the pseudospectral must be: with less collocation points the pseudospectral method is faster and may still perform better, or the other way around, enforcing a higher accuracy of the MILP model will increase its computation time. Wang and Goverde (2016a) and Ye and Liu (2016, 2017) observed strong oscillating behaviour in both the control and state for the singular solution corresponding to a nonsmooth approximation of the cruising regime. Zhong, Lin, Loxton, and Teo (2019) modelled the optimal train control problem as an optimal switching control problem. They divide the track into a finite number of segments of constant gradient and speed limit similar to a multi-phase optimal control problem, and then use control parametrization and time-scaling to obtain an approximate finite-dimensional nonlinear programming problem. The traction and braking control are approximated as piecewise constant functions where the switching times and jumps are the decision variables. They compared their algorithm with a pseudospectral method using GPOPS. The computation times are comparable and their solution does not show singular control oscillations, although the traction and braking control are now approximate step functions. Instead, the pseudospectral method shows accurate continuous traction control except for the singular control oscillations. Chen and Biegler (2016) proposed a nested direct transcription optimization method for solving singular optimal control problems. The nonlinear programming problem resulting from the direct transcription is decomposed into an inner and outer problem. In the inner problem moving finite elements are used to find accurate switching times and additional monotonic control constraints on each finite element guarantee a low-order control. ‘Pseudomultipliers are introduced that reconstruct the necessary optimality conditions for the singular optimal control in the outer problem.

In this paper, we propose a computational evaluation framework where the PMP is applied to verify, validate and improve pseudospectral solutions to optimal train control problems. It is based on the costate approximations that the pseudospectral methods provide next to the controls and states. This can be exploited to analyse the pseudospectral results using the necessary optimality conditions of the original continuous optimal control problem. It is shown that there are inconsistencies due to the discretization of the continuous problem. The solutions can however be corrected using analytical results from the PMP analysis in a postprocessing step resulting in feasible continuous solutions satisfying the optimality conditions. We consider both the MTTC and EETC problems as examples, and exclude regenerative braking. However, the proposed approach can be applied to any optimal train control problem formulation. The paper thereby gives the following contributions to the literature:

1. A computational framework connects the pseudospectral method with Pontryagin's Maximum Principle.
2. The framework is applied to compute, validate and improve solutions to optimal train control problems.
3. Structured experiments illustrate the impact of stop distances, varying speed limits and gradients, and running time supplements.
4. Convergence issues are identified for discontinuous state constraints (speed limits) and dynamic equations (gradients) with big jumps.
5. Known computational issues of the pseudospectral method regarding singular solutions are solved by a hybrid approach using Pontryagin's Maximum Principle.

The paper is structured as follows. Section 2 defines the MTTC and EETC problems and derives their necessary optimality conditions by application of the PMP. This reflects the traditional indirect optimal train control solution approach. Then, the numerical pseudospectral method is introduced in Section 3. Section 4 proposes the computational evaluation framework that combines PMP and the pseudospectral method to verify and validate the solutions obtained from the pseudospectral methods. Section 5 applies the computational framework to the MTTC and EETC problems for various structured scenarios and identifies the inconstistencies of the approximated pseudospectral solutions with the PMP. Section 6 then discusses the numerical challenges of the discretized pseudospectral solutions and proposes a postprocessing step to obtain feasible continuous solutions by exploiting knowledge from the PMP. Section 7 ends the paper with the conclusions.

2. The optimal train control problems and necessary optimality conditions

This section gives the problem formulations of the optimal control problem for both the EETC and MTTC, and derives necessary conditions for optimality by application of the PMP (Lewis, Vrabie, & Syrmos, 2012; Pontryagin et al., 1962; Ross, 2015). First, the EETC problem is discussed and afterwards the MTTC problem. The problem formulations consider distance as independent variable rather than time, because discontinuity points associated with changes in gradients and speed limit are naturally given in terms of distance (Scheepmaker & Goverde, 2015; Scheepmaker et al., 2017; Wang & Goverde, 2016a). To focus on the essence of the optimal control problems, we do not consider the effect of regenerative braking and we model the train without loss of generality as a point mass (Brünger & Dahlhaus, 2014; Howlett & Pudney, 1995). The application of the PMP to optimal train control problems is not new, but is included here to make the paper self-contained. The derived equations from the necessary optimality conditions will be used in the rest of the paper. Other papers that also give in-depth derivations of the optimality conditions by applying PMP consider slightly different problem formulations, such as energy and time as state variables (Khmelnitsky, 2000), normalized control variables (Liu & Golovitcher, 2003), and a different objective function including regenerative braking (Albrecht et al., 2016a; 2016b). Howlett and Pudney (1995) considers various problem formulations with extensive derivations. Obviously the optimal control structure is always the same but the algebraic expressions of the optimality conditions depend on the specific problem formulation, and these are used later in the paper in connection to the pseudospectral solutions.

In our problem formulation we model a typical maximum traction force as function of speed using a pure control constraint and a mixed state-control constraint. In addition, static speed limits are

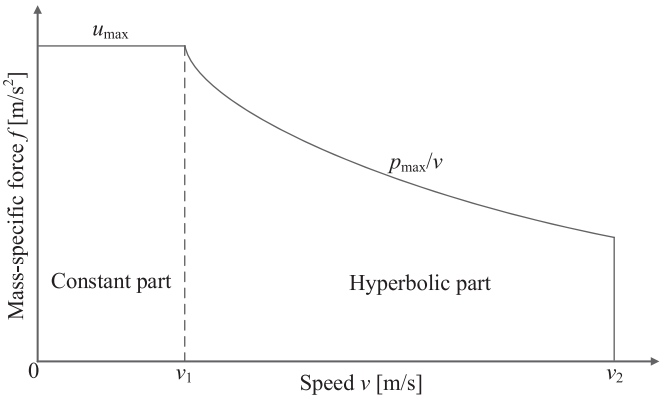

Fig. 1. Typical traction force-speed diagram with a constant and hyperbolic part.

formulated as pure state constraints. The problem formulation thus includes all types of algebraic path constraints (state, control, and mixed constraints), which will be used later to analyse their impact on the pseudospectral solution method.

2.1. Energy-efficient train control

We consider the problem of finding the optimal control for a train run between two stops in a given scheduled time T such that the total traction energy (J [m^2/s^2]) is minimized. This can be formulated as the following constrained optimal control problem:

$$\text{Minimize } J = \int_{s_0}^{s_f} u^+(s)ds, \tag{1}$$

subject to the constraints

$$\dot{t}(s) = \frac{1}{v(s)} \tag{2}$$

$$\dot{v}(s) = \frac{u(s) - r(v) - g(s)}{v(s)} \tag{3}$$

$$u^+(s)v(s) \le p_{max} \tag{4}$$

$$0 \le v(s) \le v_{max}(s) \tag{5}$$

$$-u_{min} \le u(s) \le u_{max} \tag{6}$$

$$t(s_0) = 0, t(s_f) = T, v(s_0) = 0, v(s_f) = 0, s_0 = 0, s_f = S, \tag{7}$$

where the independent variable is distance s [m], the state variables are time t [s] and speed v [m/s], $\dot{t} = dt/ds$ and $\dot{v} = dv/ds$ denote the derivatives of the state variables with respect to the independent variable s, and the control variable u [m/s^2] is the mass-specific applied force $u(s) = F(s)/(\rho m)$, i.e., the applied force divided by total mass including a rotating-mass factor ρ [-], consisting of a traction part $u^+(s)$ and a braking part $u^-(s)$. The control is bounded between a maximum specific braking rate $-u_{min} = -F_{min}/(\rho m)$ and a maximum specific traction force $u_{max} = F_{max}/(\rho m)$. Moreover, the mass-specific traction power $p = u^+v$ [m^2/s^3] is limited by a maximum mass-specific power p_{max}. Hence, the maximum traction force is a function of speed consisting of a constant and hyperbolic part which is illustrated in the traction-speed diagram of Fig. 1. It follows that the control variable is bounded by $u \in U(v) = [-u_{min}, \min(u_{max}, p_{max}/v)]$. We assumed a constant maximum braking force as this was the only braking data available. Note that based on the definition of the control variable, traction and braking control cannot be used at the same time. We use the notation $u^+(s) = \max(u(s), 0) \ge 0$ and $u^-(s) = \min(u(s), 0) \le 0$ so that $u(s) = u^+(s) + u^-(s)$. The resistance forces consist of a mass-specific train resistance $r(v) = R(v)/(\rho m)$ [m/s^2] and a mass-specific line resistance $g(s) = G(s)/(\rho m)$ [m/s^2]. The train resistance is defined by the Davis equation $r(v) = r_0 + r_1 v + r_2 v^2$, with non-negative coefficients $r_0, r_1 \ge 0$ and $r_2 > 0$ (Davis, 1926). The line resistance $g(s)$ is defined as the specific gravity force due to track gradients. It is assumed that tracks have piecewise constant gradients. Note that on uphill slopes $g(s) > 0$ and on downhill slopes $g(s) < 0$. In addition, line resistance may have an additional nonnegative term due to curvature which also depends on the location of the curves. Finally, the speed is bounded above by a speed limit $v_{max}(s)$, which is assumed piecewise constant.

Next, we derive the necessary optimality conditions using the PMP. For this, define the Hamiltonian H [m/s^2] as

$$H(t, v, \lambda_1, \lambda_2, u, s) = -u^+ + \frac{\lambda_1}{v} + \frac{\lambda_2(u - r(v) - g(s))}{v}, \tag{8}$$

where λ_1 [m^2/s^3] and λ_2 [m/s] are the costate variables, which are also functions of the independent variable s. Note that the Hamiltonian depends on distance s via the line resistance $g(s)$, which is a piecewise constant function of distance. For sections of constant gradient the Hamiltonian is independent on s and thus constant over distance, i.e. $\partial H/\partial s = 0$, although the Hamiltonian may have jumps at the points where the gradient changes and thus become piecewise constant. Also jumps in the costate may lead to jumps in the Hamiltonian. To take into account the additional (mixed) state and control path constraints (4)–(6), we define the augmented Hamiltonian $\bar{H}$ [m/s^2] as the Lagrangian of the Hamiltonian as follows:

$$\begin{aligned}\bar{H}(t, v, \lambda_1, \lambda_2, \mu, u, s) =& H + \mu_1(u_{max} - u) + \mu_2(u + u_{min}) \\ &+ \mu_3(p_{max} - u^+v) + \mu_4(v_{max} - v),\end{aligned} \tag{9}$$

where μ_1 [-], μ_2 [-], μ_3 [s/m] and μ_4 [1/s] are the Lagrange multipliers, with $\mu_i \ge 0$ $(i = 1, \ldots, 4)$. The costates λ_1 and λ_2 satisfy the differential equations $\dot{\lambda}_1(s) = -\partial\bar{H}/\partial t$ and $\dot{\lambda}_2(s) = -\partial\bar{H}/\partial v$, which gives

$$\dot{\lambda}_1(s) = 0 \tag{10}$$

$$\dot{\lambda}_2(s) = \frac{\lambda_1 + v\lambda_2 r'(v) + \lambda_2(u - r(v) - g(s))}{v^2} + \mu_3 u^+ + \mu_4. \tag{11}$$

From the first equation follows immediately that $\lambda_1(s) \equiv \lambda_1$ is constant. Note that these dynamic equations (10),(11) together with the state equations (2),(3) define a BVP with four differential equations in the states and costates and four fixed (begin and final) endpoints for the states given in (7). Therefore, the endpoints for the costates are free.

According to the PMP the optimal control maximizes the Hamiltonian (Pontryagin et al., 1962). Therefore, the optimal control is defined by

$$\hat{u}(s) = \arg\max_{u \in U} H(\hat{t}(s), \hat{v}(s), \hat{\lambda}_1(s), \hat{\lambda}_2(s), u, s), \tag{12}$$

where $(\hat{t}, \hat{v})$ and $(\hat{\lambda}_1, \hat{\lambda}_2)$ are the state and costate trajectories corresponding to the optimal control trajectory $\hat{u}$. Moreover, the maximized Hamiltonian is independent on u and thus constant when it is also independent on s. In our case the Hamiltonian depends on distance s through the piecewise constant gradient $g(s)$, and also the control, state and costate trajectories may depend on s through the state constraint (5) if the speed limit $v_{max}(s)$ is not constant. Hence, on intervals with constant $g(s) \equiv g_r$ and $v_{max}(s) \equiv v_{max,r}$, there exists a constant φ such that

$$H(\hat{t}(s), \hat{v}(s), \hat{\lambda}_1(s), \hat{\lambda}_2(s), \hat{u}(s)) = \varphi. \tag{13}$$

Note that the Hamiltonian (8) is piecewise linear in the control u, and can be split into traction and braking parts as

$$H(t,v,\lambda_1,\lambda_2,u,s)=\begin{cases}(\frac{\lambda_2}{v}-1)u+\frac{\lambda_1-\lambda_2 r(v)-\lambda_2 g(s)}{v} & \text{if } u\geq 0\\ \frac{\lambda_2}{v}u+\frac{\lambda_1-\lambda_2 r(v)-\lambda_2 g(s)}{v} & \text{if } u<0.\end{cases} \tag{14}$$

The optimal control u that maximizes the Hamiltonian thus depends on the sign of u and the relative value of speed v and the costate λ_2, which gives five cases:

1. If $\lambda_2 > v$ then u must be maximal.
2. If $\lambda_2 = v$ then $u \in [0, u_{\max}]$ is undetermined (1st singular solution).
3. If $0 < \lambda_2 < v$ then $u = 0$.
4. If $\lambda_2 = 0$ then $[-u_{\min}, 0]$ is undetermined (2nd singular solution).
5. If $\lambda_2 < 0$ then u must be minimal.

To find a full characterization of the optimal control and associated state and costate trajectories, we apply the Karush–Kuhn–Tucker (KKT) conditions on the augmented Hamiltonian (9). These conditions consist of the stationary and complementary conditions associated to $\bar{H}$ (Bertsekas, 1999). The complementary slackness conditions on the path constraints are $\mu_i \geq 0$, $i = 1, \ldots, 4$, and

$$\begin{aligned}&\mu_1(u_{\max}-u)=0, \quad \mu_2(u+u_{\min})=0,\\ &\mu_3(p_{\max}-u^+v)=0, \quad \mu_4(v_{\max}-v)=0.\end{aligned} \tag{15}$$

Note that $\mu_2 = 0$ if $u \geq 0$, and likewise $\mu_1 = \mu_3 = 0$ if $u < 0$. The stationary condition states that $\partial\bar{H}/\partial u = 0$, where the partial derivative of the augmented Hamiltonian can be split in two according to the sign of the control variable and does not exist at $u = 0$ due to the discontinuity of H at $u = 0$. The stationary condition can then be derived as

$$\begin{cases}-1+\frac{\lambda_2}{v}-\mu_1-\mu_3 v=0 & \text{if } u>0\\ \frac{\lambda_2}{v}+\mu_2=0 & \text{if } u<0,\end{cases} \tag{16}$$

and it is undefined for $u = 0$.

Based on the PMP and KKT conditions we can now characterize each of the five driving regimes depending on the costate λ_2 as follows:

1. *Maximum acceleration (MA).* The first case is $\lambda_2 > v$ with u maximal. In this case, the stationary condition (16) gives $\mu_1 + \mu_3 v = -1 + \lambda_2/v > 0$ and thus $\mu_1 > 0$ or $\mu_3 > 0$. Therefore, either $u = u_{\max}$ or $u = p_{\max}/v$. Note that $u = u_{\max}$ and $u = p_{\max}/v$ cannot hold both at the same time except at one point $v_1 = p_{\max}/u_{\max}$ since the hyperbolic function is decreasing in v. Hence, $u = u_{\max}(v) = \min(u_{\max}, p_{\max}/v)$, which specifies maximum traction, and μ_1 and μ_3 can be expressed in terms of v and λ_2 using the stationary condition (16) with either $\mu_3 = 0$ or $\mu_1 = 0$. In addition, during this maximal acceleration regime the speed bound $v = v_{\max}$ cannot be maintained except at a single point, so $\mu_4 = 0$. This completely defines the costate equation(11).
2. *Cruising by partial traction (CR1).* The second case is the singular solution with $\lambda_2 = v$ and $u \in [0, u_{\max}]$. In this case, the stationary condition (16) reads $\mu_1 + \mu_3 v = 0$ and therefore $\mu_1 = \mu_3 = 0$. Moreover, since $\lambda_2(s) = v(s)$ on a nontrivial interval, we must have $\dot{\lambda}_2(s) = \dot{v}(s)$. From (3) and (11), we can then derive $v^2(\mu_4 + r'(v)) + \lambda_1 = 0$. This equation has a unique solution v_c which can be proved as follows. If we view the left-hand side as a function $h(v)$, and recalling that $r(v)$ is a quadratic function of speed with in particular $r'(v) = r_1 + 2r_2 v \geq 0$, $r''(v) = 2r_2 > 0$, and third derivative $r^{(3)}(v) = 0$, then $h'(v) = 2v\mu_4 + 2vr'(v) + v^2 r''(v) \geq 0$ and $h''(v) = 2\mu_4 + 2r'(v) + 4vr''(v) > 0$. Hence, $h(v)$ is an increasing convex function, which has a unique root $h(v_c) = 0$ for $v_c > 0$, if any. Now assume that $v_c < v_{\max}$ and so $\mu_4 = 0$. Then the cruising speed must satisfy
$$v_c^2 r'(v_c) + \lambda_1 = 0, \tag{17}$$
which has a unique solution v_c if $\lambda_1 < 0$. Moreover, the maximized Hamiltonian value condition (13) applied to this case for an interval with constant $g(s) \equiv g$ then gives
$$v_c r'(v_c) + r(v_c) + g + \varphi = 0 \tag{18}$$
with φ piecewise constant negative in congruence to $g(s)$, and thus the Hamiltonian is negative ($H = \varphi < 0$). If on the other hand $v(s) = v_{\max}$ on a nontrivial interval then μ_4 is determined as $\mu_4 = -\lambda_1/v_{\max}^2 - r'(v_{\max})$ which gives cruising at the maximal speed $v_{\max}$ as the unique solution. Note that also in this case $\lambda_1 < 0$ must hold. In conclusion, this case corresponds to a constant cruising speed $v(s) = \min(v_c, v_{\max})$ with v_c the solution to (17). This implies that $\dot{v}(s) = 0$ and thus $u(s) = r(v) + g(s)$, i.e., the traction force equals the total resistance force to maintain the cruising speed. This case thus reduces to finding a solution v_c and λ_1 to (17).
3. *Coasting (CO).* In the third case $\lambda_2 \in (0, v)$ and $u = 0$. Since $u(s) = 0$ the complementary slackness conditions give $\mu_1 = \mu_2 = \mu_3 = 0$ and also $\mu_4 = 0$ since speed changes without traction effort due to the resistance, and, therefore, will not stay at the speed limit for a nontrivial interval. This completely defines the costate equation (11). Note that the rare case of $r(v) = -g(s)$ for some interval reduces to case 2 or 4 with cruising at zero traction $u(s) = 0$.
4. *Cruising by partial braking (CR2).* The fourth case is the second singular solution with $\lambda_2 = 0$ and $u \in [-u_{\min}, 0]$. The stationary condition (16) now gives $\mu_2 = 0$, and also $\mu_1 = \mu_3 = 0$ since $u < 0$. For the nontrival case that $\lambda_2 = 0$ on some interval, we must have $\dot{\lambda}_2 = 0$. Thus from (11) we get $\lambda_1/v^2 + \mu_4 = 0$. If $\mu_4 = 0$ then also $\lambda_1 = 0$ and by (8) also $\varphi = 0$. However, both λ_1 and φ should be negative otherwise other regimes would be prohibited, so $\mu_4 > 0$. Therefore, $v(s) = v_{\max}$ with $\mu_4 = -\lambda_1/v_{\max}^2$. Hence, partial braking is optimal only when cruising at the speed limit with $u(s) = r(v_{\max}) + g(s)$, which can occur only on downward slopes with $-u_{\min} - r(v_{\max}) < g(s) < -r(v_{\max})$.
5. *Maximum braking (MB)* In the final case $\lambda_2 < 0$ with u minimal. The stationary condition (16) now gives $\mu_2 = -\lambda_2/v > 0$ and, therefore, $u = -u_{\min}$. The special case that $v = v_{\max}$ over a nontrivial interval, i.e., $g(s) = -u_{\min} - r(v_{\max})$ is part of the fourth case (with full braking). Hence, the costate equation (11) applies with $\mu_4 = 0$, and also $\mu_3 = 0$ since $u < 0$.

The above analysis thus leads to the following optimal control structure:

$$\hat{u}(s)=\begin{cases}u_{\max}(v(s)) & \text{if } \lambda_2(s)>v(s) & \text{(MA)}\\ r(v(s))+g(s)\in[0,u_{\max}] & \text{if } \lambda_2(s)=v(s) & \text{(CR1)}\\ 0 & 0<\lambda_2(s)<v(s) & \text{(CO)}\\ r(v_{\max}(s))+g(s)\in[-u_{\min},0] & \text{if } \lambda_2(s)=0 & \text{(CR2)}\\ -u_{\min} & \text{if } \lambda_2(s)<0 & \text{(MB).}\end{cases} \tag{19}$$

The optimal energy-efficient train control thus consists of maximum traction force $u_{\max}(v)$ for maximum acceleration (MA), partial traction to counterbalance the resistance forces and cruise at an optimal cruising speed $\min(v_c, v_{\max})$ (CR1), zero traction $u = 0$ for coasting (CO) (i.e., rolling without using the engine or brakes of the train), partial braking to cruise at the maximum speed on a sufficiently downward slope (CR2), and full braking force for maximum braking (MB). The key challenge is to find the optimal

switching points between the different driving regimes and the order of the regimes to determine the optimal driving strategy. For a simple flat track (zero gradient), no speed limit, and sufficient running time supplement, the optimal driving strategy is given by the sequence MA-CR1-CO-MB or MA-CO-MB if the speed limit cannot be reached within the given time. Note that running time supplements are the extra running times above the technical minimum running time included in the timetable to deal with variations in the running time or to recover from small delays. If the train runs punctual, they can be used for energy-efficient train driving (Scheepmaker & Goverde, 2015).

Note that the dynamic equations (2) and (3) are not defined at zero speed. In practice, we use slightly positive speeds at the end points to avoid the divide by zero issue and correct the total distance and time accordingly. From Pontryagins Maximum Principle it is known that a train starts with maximum acceleration and ends with maximum braking, so this control can be simply extrapolated from/to zero. For our case studies the error is less than 25 cm or 1 s at the start and end point. In this paper, we ignored this error, which is well within the practical stopping accuracy. Likewise, for the minimum-time train control of the next subsection.

2.2. Minimum-time train control

The aim of the MTTC problem is to minimize the total running time of a train, so the train should arrive as early as possible at the next station at the minimized time $t(s_f)$. This objective J [s] can be formulated as

$$\text{Minimize } J = t(s_f), \tag{20}$$

subject to the constraints (2)–(6) and the endpoint conditions

$$t(s_0) = 0, v(s_0) = 0, v(s_f) = 0, s_0 = 0, s_f = S, \tag{21}$$

while the final time $t(s_f)$ is free.

The PMP analysis is similar to Section 2.1 and, therefore, we focus on the main points here only. The Hamiltonian H [s/m] and augmented Hamiltonian $\bar{H}$ [s/m] are defined as

$$H(t, v, \lambda_1, \lambda_2, u, s) = \frac{\lambda_1}{v} + \frac{\lambda_2(u - r(v) - g(s))}{v}, \tag{22}$$

$$\bar{H}(t, v, \lambda_1, \lambda_2, \mu, u, s) = H + \mu_1(u_{\max} - u) + \mu_2(u + u_{\min}) + \mu_3(p_{\max} - u^+ v) + \mu_4(v_{\max} - v), \tag{23}$$

where λ_1 [-] and λ_2 [s^2/m] are the costate functions of the independent variable s, and μ_1 [s^3/m^2], μ_2 [s^3/m^2], μ_3 [s^4/m^3] and μ_4 [s^2/m^2], $\mu_i \geq 0$ $(i = 1 \ldots, 4)$, are the Lagrange multipliers associated to the path constraints. Again the Hamiltonian is piecewise constant with possible jumps at the distances s where the gradient $g(s)$ or speed limit $v_{\max}(s)$ change, and (13) also holds for this Hamiltonian. The costates λ_1 and λ_2 satisfy the differential equations $\dot{\lambda}_1(s) = -\partial\bar{H}/\partial t$ and $\dot{\lambda}_2(s) = -\partial\bar{H}/\partial v$, which gives $\dot{\lambda}_1 = 0$ and thus $\lambda_1(s) \equiv \lambda_1$ as in the EETC case, and

$$\dot{\lambda}_2(s) = \frac{\lambda_1 + v\lambda_2 r'(v) + \lambda_2(u - r(v) - g(s))}{v^2} + \mu_3 u^+ + \mu_4. \tag{24}$$

However, different from the EETC case, the final conditions for the states are not all fixed now, since the final time $t(s_f)$ is free. The transversality conditions now specify a fixed value for the final time of the associated costate λ_1. This can be found by considering the endpoint Lagrangian $\bar{E}$ [s] defined as

$$\bar{E}(t(s_f), v(s_f), s_f) = -t(s_f) + \gamma_1 v(s_f) + \gamma_2(s_f - S), \tag{25}$$

where γ_1 and γ_2 are Lagrange multipliers, with the complementary slackness conditions $\gamma_1 v(s_f) = 0$ and $\gamma_2(s_f - S) = 0$. The transversality conditions state that $\lambda_1(s_f) = \partial\bar{E}/\partial t(s_f) = -1$ and therefore we now obtain $\lambda_1 \equiv -1$. Likewise, $\lambda_2(s_f) = \partial\bar{E}/\partial v(s_f) = \gamma_1$ and for the value of the maximized Hamiltonian at the endpoint s_f, we get $H[s_f] = -\partial\bar{E}/\partial s_f = -\gamma_2$, which however does not give any new information.

The Hamiltonian is again linear in the control u with coefficient λ_2/v. The optimal control u that maximizes this Hamiltonian therefore depends on the sign of the costate λ_2, and must be maximal for $\lambda_2 > 0$, is undetermined for $\lambda_2 = 0$, and is minimal for $\lambda_2 < 0$. The KKT conditions for the augmented Hamiltonian (23) give the same complementary slackness conditions (15) on the path constraints with $\mu_i \geq 0$, $i = 1, \ldots, 4$, as the EETC case, while the stationary conditions with $\partial\bar{H}/\partial u = 0$ now become

$$\begin{cases} \frac{\lambda_2}{v} - \mu_1 - \mu_3 v = 0 & \text{if } u > 0 \\ \frac{\lambda_2}{v} + \mu_2 = 0 & \text{if } u < 0, \end{cases} \tag{26}$$

and is undefined for $u = 0$. Like the EETC case, we have $\mu_2 = 0$ for $u \geq 0$, and $\mu_1 = \mu_3 = 0$ for $u < 0$. We can now characterize the three driving regimes depending on the sign of the costate λ_2 using the results of the PMP and KKT conditions.

1. *Maximum acceleration (MA).* In the first case, $\lambda_2 > 0$ and $u > 0$. Then (26) gives $\mu_1 + \mu_3 v = \frac{\lambda_2}{v} > 0$ and thus $\mu_1 > 0$ or $\mu_3 > 0$. So either $u = u_{\max}$ or $u = p_{\max}/v$, and therefore $u = u_{\max}(v) = \min(u_{\max}, p_{\max}/v)$, which specifies that maximum traction needs to be applied. In addition, during a maximal acceleration regime the speed bound $v = v_{\max}$ cannot be maintained except at a single point, so $\mu_4 = 0$. The multiplier μ_3 is either zero or equal to $\mu_3 = \lambda_2/v^2$ using (26) with $\mu_1 = 0$. This completely defines the costate equation (24).
2. *Cruising (CR).* The second case is the singular solution with $\lambda_2 = 0$ and $u \in [-u_{\min}, u_{\max}]$. With $\lambda_2 = 0$, (26) specifies $\mu_1 = \mu_2 = \mu_3 = 0$ for both $u > 0$ and $u < 0$. The Hamiltonian (22) now becomes $H = \lambda_1/v(s) = -1/v(s)$. Moreover, $\dot{\lambda}_2 = 0$ on a nontrivial interval and then $\mu_4 = 1/v^2$ using (24). In particular, this implies that $\mu_4 > 0$ and therefore, $v = v_{\max}$. Hence, this case corresponds to cruising at the maximum speed $v_{\max}$ with control $u(s) = r(v_{\max}) + g(s) \in [-u_{\min}, u_{\max}]$ to maintain the maximum speed. The control can be anything from full to partial traction or braking depending on the value of the resistance and gradient at $v_{\max}$. Moreover, the Hamiltonian value is fixed at $H = -1/v_{\max} < 0$.
3. *Maximum braking (MB)* In the final case, $\lambda_2 < 0$ and $u < 0$. Now (26) gives $\mu_2 = -\lambda_2/v > 0$ and therefore $u = -u_{\min}$. The special case that $v = v_{\max}$ over a nontrivial interval is part of the 2nd case (with full braking). Hence, $\mu_4 = 0$, and also $\mu_3 = 0$ since $u < 0$, which completely defines the costate equation (24). Finally, from (22) follows that $H = \varphi < 0$ since both costates are smaller than zero.

The above analysis thus leads to the optimal control structure

$$\hat{u}(s) = \begin{cases} u_{\max}(v(s)) & \text{if } \lambda_2(s) > 0 \quad \text{(MA)} \\ r(v_{\max}(s)) + g(s) & \text{if } \lambda_2(s) = 0 \quad \text{(CR)} \\ -u_{\min} & \text{if } \lambda_2(s) < 0 \quad \text{(MB)}. \end{cases} \tag{27}$$

So, the traction and braking efforts are as high as possible to keep the train at its maximum speeds as long as possible, which indeed generates the minimal running time. No time is wasted on coasting in this case.

3. Pseudospectral method

In this section, the train trajectory optimization is formulated as a multiple-phase optimal control problem (Betts, 2010; Darby et al., 2011a) and then discretized according to the Radau pseudospectral method (Garg, 2011; Garg et al., 2009; Rao et al., 2010).

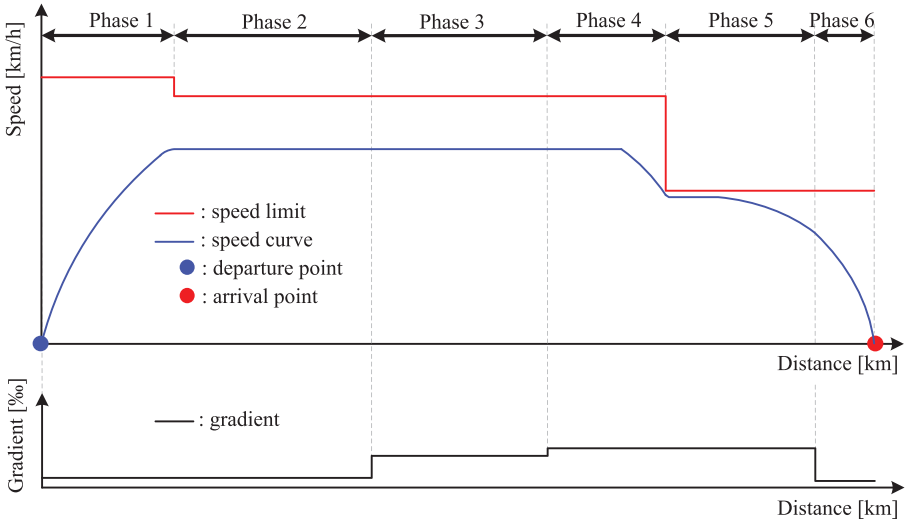

Fig. 2. Illustration of phases in the multiple-phase optimal control problem.

The Radau pseudospecral method is implemented in the MATLAB toolbox GPOPS, which we use as solver for the MTTC and EETC problems.

A multiple-phase optimal control problem is one where the trajectory consists of a collection of phases. Simply speaking, a phase is any segment of the complete trajectory. In general, any particular phase of an optimal control problem has a cost functional, dynamic constraints, path constraints, and boundary conditions. However, the models used to quantitatively describe the trajectory may be different in different phases of the trajectory. The complete trajectory is then obtained by properly linking adjacent phases via linkage conditions. Similarly, the total cost functional is the sum of the cost functionals over each phase. The optimal trajectory is then found by minimizing the total cost functional subject to the constraints within each phase and the linkage constraints connecting adjacent phases.

The multiple-phase optimal control problem described in the previous section is now transcribed to a discrete NLP via an extension of the single-phase Radau pseudospectral method. Pseudospectral methods are orthogonal collocation methods using global poynomials. A collocation method can be used for the numerical solution of ordinary differential equations and integral equations. The idea is to choose a finite-dimensional space of candidate solutions (here polynomials) and a number of points in the domain (called collocation points), and to select that solution which satisfies the given equation at the collocation points. The Radau pseudospectral method uses the Legendre–Gauss–Radau (LGR) points plus additionally the final point. The resulting NLP can then be solved by one of the many well developed nonlinear optimization algorithms.

3.1. Multiple-phase optimal control problem formulation

The optimal train control problems can be formulated as finite-horizon multiple-phase optimal control problems (Wang & Goverde, 2016a). The running section from the departure point to the arrival point is then divided into a finite number of segments by the critical points of speed limits, gradients and curves, and each of these segments is called a "phase". Fig. 2 gives an illustration, where the running section is divided into six phases. Within each phase, the gradient and speed limit are constant, but their values may be different over the various phases.

Assume that the running section is divided into phases $[s_0^{(r)}, s_f^{(r)}]$, with $s_0^{(r)}$ and $s_f^{(r)}$, $s_0^{(r)} < s_f^{(r)}$, denoting the initial and terminal location of phase $r \in \{1, \ldots, R\}$. The train departs from the initial point $s_0^{(1)}$ of phase 1, and arrives at the terminal point $s_f^{(R)}$ of phase R. Within phase r, the state vector consists of time and speed $x^{(r)}(s) = [t^{(r)}(s), v^{(r)}(s)]'$, and the control vector is the mass-specific applied force $u^{(r)}(s)$. Moreover, denote the boundary points of a phase r as $x_0^{(r)} = x^{(r)}(s_0^{(r)})$ and $x_f^{(r)} = x^{(r)}(s_f^{(r)})$.

The multiple-phase optimal control problem can now be formulated as minimizing the cost functional

$$\text{Minimize } J = \sum_{r=1}^{R} E^{(r)}\left(x_0^{(r)}, s_0^{(r)}, x_f^{(r)}, s_f^{(r)}\right) + \int_{s_0^{(r)}}^{s_f^{(r)}} F^{(r)}\left(x^{(r)}(s), u^{(r)}(s), s\right) ds, \tag{28}$$

subject to the dynamic constraints

$$\dot{x}^{(r)}(s) = f^{(r)}\left(x^{(r)}(s), u^{(r)}(s), s\right), \qquad r = 1, \ldots, R, \tag{29}$$

the path constraints

$$h^{(r)}\left(x^{(r)}(s), u^{(r)}(s), s\right) \geq 0, \qquad r = 1, \ldots, R, \tag{30}$$

the linkage conditions

$$l^{(r)}(x_f^{(r)}, s_f^{(r)}, x_0^{(r+1)}, s_0^{(r+1)}) = 0, \qquad r = 1, \ldots, R-1, \tag{31}$$

and the boundary (or endpoint) conditions

$$e\left(x_0^{(1)}, s_0^{(1)}, x_f^{(R)}, s_f^{(R)}\right) = 0. \tag{32}$$

Note that the path constraints are inequalities over the phases and the linkage conditions are equality constraints on the phase boundaries.

For the EETC problem the cost functional is given by $E = 0$ and $F^{(r)}(u^{(r)}) = u^{+(r)}$ with $u^{+(r)} = \max(u^{(r)}, 0)$, and for the MTTC problem these are $E^{(r)}(x_0^{(r)}, x_f^{(r)}) = t_f^{(r)} - t_0^{(r)}$ and $F^{(r)} = 0$. The dynamic and path constraints are given for both problems by (2)–(6), i.e.,

$$f^{(r)}\left(x^{(r)}(s), u^{(r)}(s), s\right) = \left[\frac{1}{v^{(r)}(s)}, \frac{u^{(r)}(s) - r^{(r)}(v^{(r)}(s)) - g^{(r)}}{v^{(r)}(s)}\right]'$$

and

$$h^{(r)}\left(x^{(r)}(s), u^{(r)}(s), s\right) = \begin{bmatrix} p_{\max} - u^{+(r)}(s)v^{(r)}(s) \\ v^{(r)}(s) \\ v_{\max}^{(r)} - v^{(r)}(s) \\ u^{(r)}(s) + u_{\min} \\ u_{\max} - u^{(r)}(s) \end{bmatrix},$$

where $g^{(r)}$ represents the line resistance due to the constant gradient within phase r, and $v^{(r)}_{\max}$ is the constant speed limit within phase r. The linkage conditions are

$$l^{(r)}(x_f^{(r)}, s_f^{(r)}, x_0^{(r+1)}, s_0^{(r+1)}) = \left[t_f^{(r)} - t_0^{(r+1)}, v_f^{(r)} - v_0^{(r+1)}, s_f^{(r)} - s_0^{(r+1)}\right]',$$

which make sure that the speed profile over the entire running section is continuous. The boundary conditions for the EETC problem are $e(x_0^{(1)}, s_0^{(1)}, x_f^{(R)}, s_f^{(R)}) = [t_0^{(1)}, v_0^{(1)}, s_0^{(1)}, t_f^{(R)} - T, v_f^{(R)}, s_f^{(R)} - S]'$ and for the MTTC problem they are $e(x_0^{(1)}, s_0^{(1)}, x_f^{(R)}, s_f^{(R)}) = [t_0^{(1)}, v_0^{(1)}, s_0^{(1)}, v_f^{(R)}, s_f^{(R)} - S]'$ with final time $t_f^{(R)}$ free.

3.2. Legendre–Gauss–Radau discretization

The Radau pseudospectral method is a pseudospectral method with the Legendre–Gauss–Radau points as collocation points (Garg, 2011; Garg et al., 2009). Pseudospectral methods are defined on the domain $[-1, 1]$. Therefore, the first step in a pseudospectral method is to map the physical domain $[s_0, s_f]$ to the computational domain $[-1, 1]$. For a multiple-phase optimal control problem, the physical variable $s \in [s_0^{(r)}, s_f^{(r)}]$ in each phase is transformed to the computational variable $\sigma^{(r)} \in [-1, 1]$ by means of the affine transformation

$$\sigma^{(r)}(s) = \frac{2s - (s_f^{(r)} + s_0^{(r)})}{s_f^{(r)} - s_0^{(r)}}. \tag{33}$$

Note this implies $d\sigma^{(r)}/ds = 2/(s_f^{(r)} - s_0^{(r)})$ for $r = 1, \ldots, R$. Conversely, we obtain $s(\sigma^{(r)}) = \frac{s_f^{(r)} + s_0^{(r)}}{2} + \frac{s_f^{(r)} - s_0^{(r)}}{2}\sigma^{(r)}$.

Next, the state vector function $x^{(r)}(s(\sigma))$ is approximated by a polynomial $X^{(r)}(\sigma)$ with a basis of global Lagrange interpolating polynomials $L_i^{(r)}(\sigma)$ over $N_r + 1$ discretization points $(\sigma_1^{(r)}, \sigma_2^{(r)}, \ldots, \sigma_{N_r+1}^{(r)})$ in each phase r in terms of σ as

$$x^{(r)}(s(\sigma)) \approx X^{(r)}(\sigma) = \sum_{i=1}^{N_r+1} X_i^{(r)} L_i^{(r)}(\sigma), \tag{34}$$

where $X_i^{(r)} = X^{(r)}(\sigma_i^{(r)})$ and

$$L_i^{(r)}(\sigma) = \prod_{j=1, j\neq i}^{N_r+1} \frac{\sigma - \sigma_j^{(r)}}{\sigma_i^{(r)} - \sigma_j^{(r)}}.$$

For the Radau pseudospectral method, the discretization points are the LGR collocation points $(\sigma_1^{(r)}, \sigma_2^{(r)}, \ldots, \sigma_{N_r}^{(r)})$ and the additional point $\sigma_{N_r+1}^{(r)} = 1$, where N_r is the number of LGR points within phase r. The LGR points are defined as the N_r zeros of the sum of the Legendre polynomials of degree N_r and $N_r - 1$. It includes the initial point $\sigma_1^{(r)} = -1$, but not the final point. Therefore, for the state approximation at the final point, the additional point $\sigma_{N_r+1}^{(r)} = 1$ is added. Note that $L_i^{(r)}(\sigma_k^{(r)})$ satisfy the Kronecker delta condition $L_i^{(r)}(\sigma_k^{(r)}) = \delta_{ik}$ for $k = 1, \ldots, N_r + 1$, with $\delta_{ik} = 1$ if $i = k$ and $\delta_{ik} = 0$ if $i \neq k$. Hence, at the discretization points $\sigma_k^{(r)}$, we have $x^{(r)}(s(\sigma_k^{(r)})) = X_k^{(r)}$. Moreover, we define $U_k^{(r)}$ as the approximation of the control at each point $\sigma_k^{(r)}$, for $k = 1, \ldots, N_r$.

The derivative of $x^{(r)}(s)$ is approximated by the derivative of the polynomial approximation, which gives in each phase r,

$$\dot{x}^{(r)}(s(\sigma)) \approx \dot{X}^{(r)}(\sigma) = \sum_{i=1}^{N_r+1} X_i^{(r)} \dot{L}_i^{(r)}(\sigma).$$

Now define the non-square $N_r \times (N_r + 1)$ Radau pseudospectral differentiation matrix in phase r,

$$D_{ki}^{(r)} = \dot{L}_i^{(r)}(\sigma_k^{(r)}), \qquad k = 1, \ldots, N_r, i = 1, \ldots, N_r + 1,$$

with the derivatives of $L_i^{(r)}(\sigma)$ evaluated at the LGR points $\sigma = \sigma_k^{(r)}$. Then the dynamic constraint (29) is approximated at the LGR points $\sigma_k^{(r)}$ by the algebraic equations

$$\sum_{i=1}^{N_r+1} D_{ki}^{(r)} X_i^{(r)} - \frac{s_f^{(r)} - s_0^{(r)}}{2} f^{(r)}\left(X_k^{(r)}, U_k^{(r)}, \sigma_k^{(r)}\right) = 0, \tag{35}$$

for $k = 1, \ldots, N_r, r = 1, \ldots, R$.

The cost function is approximated using the LGR quadrature as

$$J = \sum_{r=1}^{R} \left(E^{(r)}\left(X_1^{(r)}, X_{N_r+1}^{(r)}\right) + \sum_{k=1}^{N_r} \frac{s_f^{(r)} - s_0^{(r)}}{2} w_k^{(r)} F_k^{(r)} \right), \tag{36}$$

where $F_k^{(r)} = F^{(r)}(X_k^{(r)}, U_k^{(r)}, \sigma_k^{(r)})$ is the cost evaluated at the kth collocation point in phase r, $w_k^{(r)}$ is the quadrature weight associated with the kth LGR collocation point of phase r, and for the MTTC problem $E^{(r)}(X_1^{(r)}, X_{N_r+1}^{(r)}) = X_{N_r+1,1}^{(r)} - X_{1,1}^{(r)}$ with $X_{k,1}^{(r)}$ the 1st (time) component of $X_k^{(r)} = (T_k^{(r)}, V_k^{(r)})$. Note that for the optimal train control problem $X^{(r)}$ is a matrix with rows $X_i^{(r)}$ corresponding to the discretization points and thus the columns correspond to the approximation vectors of the state variables.

Furthermore, the path constraints, linkage conditions and boundary conditions are discretized at the LGR points as

$$h^{(r)}\left(X_k^{(r)}, U_k^{(r)}, \sigma_k^{(r)}\right) \geq 0, \quad k = 1, \ldots, N_r, r = 1, \ldots, R, \tag{37}$$

$$l^{(r)}(X_{N_r+1}^{(r)}, \sigma_{N_r+1}^{(r)}, X_1^{(r+1)}, \sigma_1^{(r+1)}) = 0, \quad r = 1, \ldots, R-1, \tag{38}$$

$$e\left(X_1^{(1)}, \sigma_1^{(1)}, X_{N_r+1}^{(R)}, \sigma_{N_r+1}^{(R)}\right) = 0. \tag{39}$$

The multiple-phase optimal control problem is now converted to the discrete nonlinear programming problem to minimize (36) subject to (35), (37)–(39), where the decision variables are $(X_1^{(r)}, \ldots, X_{N_r+1}^{(r)})$ and $(U_1^{(r)}, \ldots, U_{N_r}^{(r)})$, $r = 1, \ldots, R$. Necessary optimality conditions for the NLP problem (35)–(39) can be obtained by the associated KKT conditions including the corresponding Lagrangian multipliers. Let $\Lambda^{(r)}$ denote the Lagrange multipliers associated to the discretized dynamic equation (35), i.e., $\Lambda^{(r)}$ is an $N_r \times 2$ matrix whose columns correspond to the Lagrange multiplier vectors of the two columns in $X^{(r)}$ representing the state approximations. Then the approximation of the costate vector $\lambda^{(r)}$ in the discretization point $\sigma_k^{(r)}$ is given as (Garg et al., 2011)

$$\lambda^{(r)}(s(\sigma_k^{(r)})) = \begin{cases} \Lambda_k^{(r)}/w_k^{(r)}, & k = 1, \ldots, N_r \\ D_{N_r+1}^{\top} \Lambda^{(r)}, & k = N_r + 1, \end{cases} \tag{40}$$

where $D_{N_r+1}^{\top}$ denotes the transposed last *column* of the differentiation matrix D, which can be shown to be equal to the (transposed) negative sum of the first N_r columns of D (Garg et al., 2011).

Note that $s(\sigma_{N_r+1}^{(r)}) = s(\sigma_1^{(r+1)})$ for $r = 1, \ldots, R-1$ in the discretized multiple-phase optimal control problem using LGR collocation points, and the linking conditions take care that the associated state and control variables are consistent. Hence, the multiple-phase Radau pseudospectral method computes in essence the state, costate and control values at $N = \sum_{r=1}^{R} N_r$ different points, along with the state and costate values at the final point. The value of the control at the final point $U_{N_r+1}^{(R)}$ must be extrapolated separately, which is also implemented in GPOPS. Thus, the multiple-phase Radau pseudospectral method computes state, costate and control vectors of dimension $N + 1$ corresponding to the values $(X_k^{(r)}, \Lambda_k^{(r)}, U_k^{(r)})$ at the $N + 1$ successive points $s(\sigma_k^{(r)})$, $k = 1, \ldots, N_r$, $r = 1 \ldots, R$, and $s(\sigma_{N_r+1}^{(R)})$, with $s_0 = s(\sigma_1^{(1)})$ and $s_f = s(\sigma_{N_r+1}^{(R)})$. Thus, over the physical domain the discretized solution can be denoted as $\{(X_i, \Lambda_i, U_i) = (x(s_i), \lambda(s_i), u(s_i) \mid i = 0, \ldots, N\}$.

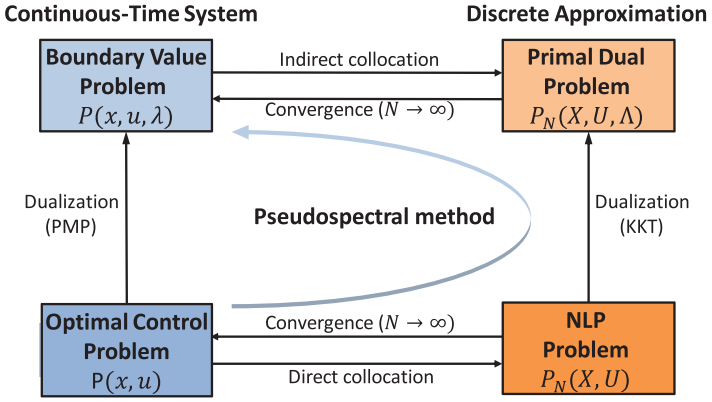

Fig. 3. Computational evaluation framework for pseudospectral optimal control with Pontryagin's Maximum Principle.

4. The pseudospectral computational evaluation framework

In this section, we propose a computational evaluation framework to verify, validate and improve pseudospectral solutions for optimal train control problems. It is based on the costate approximations that the pseudospectral methods provide next to the controls and states. These can be exploited to analyse the results using the necessary optimality conditions of the original continuous optimal control problem.

Fig. 3 illustrates the approach, which was inspired by Ross and Karpenko (2012). Here, (x, u, λ) represent the continuous state, control and costate functions, respectively, and (X, U, Λ) the discrete approximations. The indirect method is clockwise: the original optimal control problem $P(x, u)$ (lower-left box) is dualized to a BVP $P(x, u, \lambda)$ (upper-left box) with the adjoint dynamic equation of the costate λ by application of the PMP (and additional Karush–Kuhn–Tucker conditions for the path constraints). This procedure was demonstrated in Section 2. The resulting BVP $P(x, u, \lambda)$ can then be discretized by collocation of the dynamic equations and additional path and control constraints enforced on the N discretization points, which is indicated as indirect collocation in the figure (upper-horizontal arcs). However, such algorithms are very sensitive to the initial guesses of the (costate) variables. Moreover, the implicit control function, discontinuities, and the a priori unknown switching structure of the inequality path constraints make these problems very hard to solve by generic BVP solvers. Still, dedicated algorithms have been developed for the EETC problem expoiting algebraic formulae for the costate along sections with constant gradient, which led to very fast algorithms (Albrecht et al., 2016a; 2016b; Howlett et al., 2009; Liu & Golovitcher, 2003). In addition, an MTTC algorithm is straightforward with the knowledge of the optimal control structure without the need for solving the costate dynamic equations.

On the other hand, counterclockwise is the direct pseudospectral method as described in Section 3: the original optimal control problem $P(x, u)$ is transcribed into a discrete NLP problem $P_N(X, U)$ using the LGR collocation points, where N denotes the total number of collocation points (lower-right box). This NLP problem is then solved by an NLP algorithm, which gives the optimal state and control vectors that can be transformed back to approximate solutions of the continuous optimal control problem. The higher the number of discretization points N the more accurate is the approximation of the continuous functions (x, u), although the high-order global polynomials already generate good approximations for smaller grid sizes to smooth problems. The NLP algorithms also compute the associated Lagrange multipliers Λ and thus actually solve the discrete primal-dual problem $P_N(X, U, \Lambda)$ with both discrete state and costate vectors (upper-right box). This solution (X, U, Λ) can then be transformed to approximations of the solution to the continuous BVP $P(x, u, \lambda)$. Then the optimality conditions represented by $P(x, u, \lambda)$ (upper-left box) can be checked to confirm convergence of the optimal state and control of the original optimal control problem $P(x, u)$. Identified issues of the solution (x, u, λ) could be improved by increasing the number of collocation points N but this increases the computation time. Alternatively, the discrete approximation (X, U, Λ) could be corrected to a feasible continuous solution (x, u, λ) using information from the continuous PMP combined with additional computed values by the pseudospectral method, such as λ_1 and φ. This leads to a postprocessing step to the pseudospectral method, which we will explore in Section 6.

In Section 2, we derived necessary optimality conditions in terms of costate variables by application of the PMP (left-upwards arrow). This PMP analysis provides important generic knowledge that can be used to evaluate a solution found by any method. For the EETC problem the main results can be summarized as follows:

1. The Hamiltonian is negative and constant on phases with constant gradient and speed limit, $\varphi < 0$.
2. The costate associated with time is a negative constant, $\lambda_1 < 0$.
3. The optimal control structure depends on the relative values of speed v and costate λ_2 according to (19).
4. The singular solutions correspond to a constant cruising speed, with either $v = \lambda_2$ (CR1) or $\lambda_2 = 0$ (CR2).
5. In the singular solution CR1, the optimal cruising speed v_c can be determined from either λ_1 or φ using (17) or (18), while the (constrained) cruising speed is $v(s) = \min(v_c, v_{\max}(s))$.
6. In the singular solution CR2, the cruising speed is $v_{\max}$.

Likewise, for the MTTC problem the main PMP results can be summarized as:

1. The Hamiltonian is negative and constant on phases with constant gradient and speed limit, $\varphi < 0$.
2. The costate associated with time is constant and fixed, $\lambda_1 = -1$.
3. The optimal control structure depends on the sign of cospeed λ_2 according to (27).
4. The singular solutions correspond to a cruising speed $v_{\max}$ and $\lambda_2 = 0$.

In Section 5, we will apply the pseudospectral method to solve the optimal train control problems MTTC and ETTC for various scenarios and verify consistency with the necessary optimality conditions as summarized above. Then in Section 6 we use the analytical results from the PMP analysis to correct the inconsistencies in the discrete pseudospectral solutions to obtain feasible continuous solutions that satisfy the optimality conditions.

Table 1
Basic parameters of the Intercity and Sprinter trains (NS, 2017).

Characteristic	Intercity	Sprinter
Train mass m [t]	391	198
Rotating mass factor ρ [-]	1.06	1.06
Maximum traction power $P_{\max}$ [kW]	2157	1918
Maximum traction force $F_{\max}$ [kN]	214	170
Maximum braking deceleration $-u_{\min}$ [m/s^2]	−0.66	−0.8
Maximum speed limit $v_{\max}$ [km/h]	140	140
Train resistance $R(v)$ [kN] (v: [km/h])	$5.8584 + 0.0206v + 0.001v^2$	$1.3961 + 0.0145v + 0.0007v^2$

5. Computational results

This section applies the pseudospectral computational framework of Fig. 3 in a structured experimental study considering a range of operational scenarios of the optimal train control problems. In particular, we will apply the pseudospectral method to numerically compute the optimal train control problems directly, via the right loop $P(x,u) - P_N(X,U) - P_N(X,U,\Lambda) - P(x,u,\lambda)$, and then compare the computed solutions with the optimality conditions obtained via the straight arc $P(x,u) - P(x,u,\lambda)$. The purpose of these experiments is both practical and theoretical. On the practical side, we consider the following research question: how do the optimal driving strategies change depending on varying operational parameters, such as gradients, speed limits, and running time supplements? On the theoretical side, we want to answer the research question: how does the performance of the pseudospectral method depend on varying operational parameters, in terms of accuracy of the (state, costate and control) trajectories and computation time? Moreover, to improve the identified accuracy problems we consider a third research question: how can the computational framework improve the solution quality of the optimal train control problems? Section 5.1 introduces the case study and scenarios. Sections 5.2–5.5 consider the various structured experiments varying one operational parameter at a time, while Section 5.6 considers the real-life case. A solution approach to the identified issues of the discrete approximations is provided in Section 6.

5.1. Description of case study and scenarios

The case studies are based on the Dutch railway line from Utrecht Central (Ut) station to 's-Hertogenbosch (Ht) station with trains from the Netherlands Railways (NS). We consider an Intercity (IC) train running over the entire distance of 50 km from Ut-Ht, and a regional or Sprinter (SPR) train running over the last 10 km distance from the last short stop Zaltbommel (Zbm) to Ht. The IC trains use a composition of VIRM-6 rolling stock and the Sprinter trains use an SLT-6 train composition. The static parameters of the two train types are listed in Table 1. The mass-specific train related parameters in Eqs. (3)–(6), i.e., $p_{\max}$, $u_{\max}$ and $r(v)$, are computed by dividing the corresponding values in Table 1 by the total mass, including rotating mass factor. We consider various scenarios in order to investigate the effect of varying speed limits, gradients, and running time on the driving strategy and the pseudospectral convergence. The reference scenario is a level track with a single speed limit of 140 km/h (38.89 m/s), and 15% running time supplement for the EETC cases. The varying parameters of the scenarios are carefully selected to analyse a wide range of conditions and thus include gradients with steep uphill and downhill sections, slight to severe speed restrictions, and running time supplements ranging from zero to relative big allowances. The 15% running time supplement is relatively high but it refers to flat track without speed restrictions, while speed restrictions and some gradients lead to increased running times. With this choice all scenarios have feasible scheduled running times.

We consider the following scenarios with a mix of MTTC and EETC problems and IC and SPR trains:

1. Reference scenarios: fixed speed limit (140 km/h), flat track, and 15% running time supplement for EETC.
2. Varying speed limits: 10 km segment with a restricted speed of 120, 110, 100 km/h (EETC for IC only).
3. Varying (steep) gradients: 10 km segment with gradient of 10%, 5%, 0%, −5%, −10% (EETC for IC only).
4. Varying running time supplements: 0%, 2%, 5%, 10%, 15%, 20% (EETC only).
5. Real-life cases: actual varying speed limits and gradients for SPR and IC, and 10% supplement for EETC.

Note that the running time supplements only apply to the EETC problem. The results of varying speed limits and gradients are similar for the IC and SPR trains, so we only present the IC trains. In total, we analyse 25 scenarios. From a theoretical perspective, the reference scenarios focus on the main optimal train control problem with smooth state, control and mixed state-control path constraints. The 2nd class of scenarios considers discontinuous state constraints and the 3rd class of scenarios considers discontinuities in the dynamic equation. The 4th class of scenarios focuses on the interaction between regimes when more or less time is available for energy-efficient driving. The real-life cases combine the first three scenarios with the real varying gradient profile and speed limits.

We use the MATLAB toolbox GPOPS version 4.1 as the pseudospectral solver (Rao et al., 2010). GPOPS approximates the states, controls, and costates as functions of distance, as well as the Hamiltonian values, and thus closes the numerical solution loop via the discrete approximation from $P(x,u)$ to $P(x,u,\lambda)$. We aimed at using the same fixed number of collocation points for each phase in all structured scenarios. By trial and error, we found $N = 200$ to be a good trade-off between computation time and solution quality. In one case (SPR MTTC), we increased this number (to $N = 235$) to get a better solution accuracy while the computation time was still fast (within 2 s). For the real-life cases, the number of phases is very high (78 and 38 for the IC and SPR, respectively), so in these scenarios we let GPOPS decide on the optimal number of collocation points per phase by exploiting the hp-adaptive pseudospectral method functionality with 1 or 3 mesh iterations, which then adjusts the number of collocation points in each phase to improve the solution quality. This resulted in 9 to 24 collocation points per phase. As initial guesses, we provide the known parameters such as the initial and final position, and the lower and upper bounds on time, speed and control.

Table 2 summarizes the results for all 25 scenarios. A first inspection of the optimality conditions confirms negative Hamiltonian values $\varphi < 0$ and a constant negative costate $\lambda_1 < 0$ in all cases. Moreover, for the MTTC scenarios $\lambda_1 = -1$. In the scenarios with the varying speed limits and/or gradients the Hamiltonian is piecewise constant over the phases, which is also as expected.

Table 2
Main pseudospectral results of the various scenarios (energy saving w.r.t. MTTC, speed limit reference scenario is 38.89 m/s).

Scenario	Fig	Trip time [s]	Energy usage [kWh]	Energy saving [%]	Max speed [m/s]	R	N	Costate λ_1	Hamiltonian φ	Comp time [s]
Reference IC MTTC	4	1340	447.21	–	38.89	1	200	−1	−0.026	0.82
Reference IC EETC	5	1541	323.98	27.56	35.12	1	200	−2.93	−0.142	21.36
Reference SPR MTTC	6	278	75.09	–	38.89	1	235	−1	−0.025	1.22
Reference SPR EETC	7	320	42.96	42.79	36.03	1	200	−6.09	−0.241	14.23
Var. speed limit IC (120 km/h)	8 a	1541	324.04	27.54	35.68	3	3×200	−3.06	−0.146	129.62
Var. speed limit IC (110 km/h)	8 b	1541	327.32	26.81	37.07	3	3×200	−3.43	−0.156, −0.161, −0.156	117.43
Var. speed limit IC (100 km/h)	8 c	1541	338.16	24.38	38.89	3	3×200	−4.75	−0.191, −0.214, −0.191	90.23
Var. gradient IC (−0.01)	9 a	1541	218.81	51.07	38.89	3	3×200	−3.47	−0.157, −0.101, −0.157	136.40
Var. gradient IC (−0.005)	9 b	1541	269.64	39.70	35.14	3	3×200	−2.93	−0.142, −0.093, −0.142	336.43
Var. gradient IC (+0.005)	9 c	1541	382.23	14.53	35.13	3	3×200	−2.93	−0.142, −0.191, −0.142	135.35
Var. gradient IC (+0.01)	9 d	1541	437.16	2.24	35.06	3	3×200	−2.91	−0.142, −0.240, −0.142	163.41
Var. running time IC (2%)	10	1367	411.84	7.91	38.89	1	200	−10.38	−0.334	10.50
Var. running time IC (5%)	10	1407	380.27	14.97	38.89	1	200	−4.55	−0.184	14.11
Var. running time IC (10%)	10	1474	352.06	21.28	37.20	1	200	−3.46	−0.157	17.68
Var. running time IC (15%)	10	1541	323.98	27.56	35.12	1	200	−2.93	−0.142	21.36
Var. running time IC (20%)	10	1608	303.05	32.24	33.32	1	200	−2.51	−0.130	17.29
Var. running time SPR (2%)	11	284	63.14	15.91	38.89	1	200	−18.96	−0.568	15.87
Var. running time SPR (5%)	11	292	56.15	25.22	38.89	1	200	−11.43	−0.375	16.89
Var. running time SPR (10%)	11	306	48.63	35.24	37.85	1	200	−7.83	−0.285	18.76
Var. running time SPR (15%)	11	320	42.96	42.79	36.03	1	200	−6.09	−0.241	14.23
Var. running time SPR (20%)	11	334	39.12	47.90	34.69	1	200	−5.08	−0.214	16.85
Real-life MTTC IC Ut-Ht	12	1619	325.68	–	38.89	78	9–24	−1	pwc < 0	57.09
Real-life EETC IC Ut-Ht	13	1781	188.79	42.03	37.89	78	9–20	−3.02	pwc < 0	202.53
Real-life MTTC SPR Zbm-Ht	14	529	65.51	–	38.89	38	9–20	−1	pwc < 0	35.42
Real-life EETC SPR Zbm-Ht	15	582	36.76	43.89	35.26	38	9–20	−4.33	pwc < 0	63.24

Legend: pwc = piecewise constant, R = Number of phases, N = Number of collocation points per phase

In the next sections, we consider each set of scenarios with the focus on the optimal control structure and the singular solutions depending on the value of the costate λ_2 relative to speed and/or zero.

5.2. Reference scenario

This subsection describes the results of the reference scenario for both the Intercity and the Sprinter train. The reference is a flat track with a fixed speed limit of 140 km/h and for the EETC case 15% running time supplement with respect to the computed minimum running time. With the reference scenario we analyse the consistency of the driving behavior of both trains with the necessary optimality conditions for the MTTC and EETC problems derived in Section 2. Table 2 summarizes the quantitative results. The resulting diagrams of the IC train can be found in Fig. 4 (MTTC) and Fig. 5 (EETC), and the results of the SPR train are shown in Fig. 6 (MTTC) and Fig. 7 (EETC).

The MTTC driving strategy for the IC leads to the fastest running time of 1340 s, but also to the highest energy consumption of 447 kWh (see Table 2). The computation time is 0.82 s for this scenario. If we have a closer look at the results of the MTTC driving strategy for the IC train of Fig. 4, we can see that the optimal driving strategy consists of maximum acceleration, cruising at the speed limit, and maximum braking. The costate λ_2 behaves according to the optimal control structure (27). During maximum acceleration, the costate variable λ_2 is bigger than zero, during the cruising phase it is equal to zero, and during maximum braking it is smaller than zero. The Hamiltonian remains constant as required, except at the endpoints where it deviates. The control, state and costate are continuous at the end points, which implies that the Hamiltonian should also be continuous. The deviations are at the level of the 3rd decimal so they can be caused by numerical errors. We observe the same errors at the endpoints of the Hamiltonian in the other scenarios, so we indeed believe that these are caused by numerical errors rather than the solutions of the state, control or costate variables. This holds for all scenarios below as well, so in the remainder we will discard these discontinuities at the endpoints.

The EETC driving strategy for the IC generates an energy saving of about 28% compared to the MTTC driving strategy. The running time exploits the full 15% running time supplement. The results are computed in 22 s. Fig. 5 shows that the Hamiltonian is constant, and the optimal driving strategy consists of maximum acceleration, cruising below the speed limit, coasting, and maximum braking, although also a partial acceleration can be observed at the end of the maximum acceleration regime where the speed profile is still slightly increasing towards the cruising speed. Also the braking regime at the end shows some oscillations rather than full braking. The Hamiltonian shows here also some more deviations, which might indicate that these oscillations are not optimal. The costate variable λ_2 behaves according to the derived control structure (19), since λ_2 is bigger than speed v during maximum acceleration, equal to the speed v during cruising, between zero and speed v during coasting, and smaller than zero during maximum braking. The costate λ_2 also equals speed during the 'partial acceleration', which thus should correspond to the singular cruising regime. The effect of partial acceleration on the total energy consumption is less than 2% compared to full maximum acceleration. The control plot shows oscillations during cruising below the speed limit, in which the partial traction is approximated by alternating between zero and some positive traction. It is known from the literature that the pseudospectral methods suffer from this oscillating behavior for singular solutions in the interior of the state inequality path constraints, i.e., cruising below the speed limit (Scheepmaker & Goverde, 2016; Wang & Goverde, 2016a; Ye & Liu, 2016), since the control solution is here not uniquely determined. Howlett and Pudney (1995) proved that an optimal cruising speed can be approximated by (short) sequences of maximum traction and coasting with the same energy consumption as cruising at the optimal cruising speed. However, the pseudospectral method switches at successive collocation points using an appro-

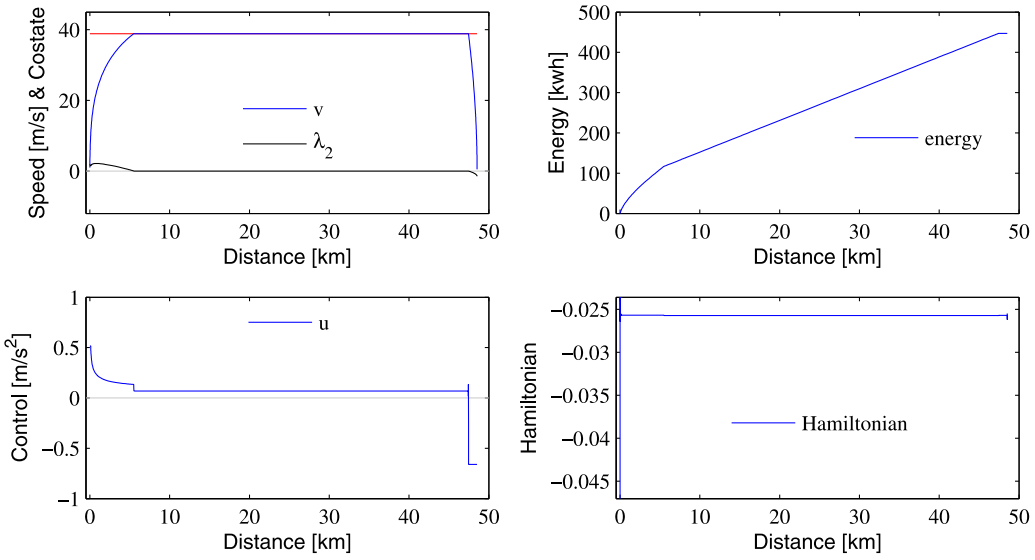

Fig. 4. Trajectories for the Intercity train with the minimum-time driving strategy.

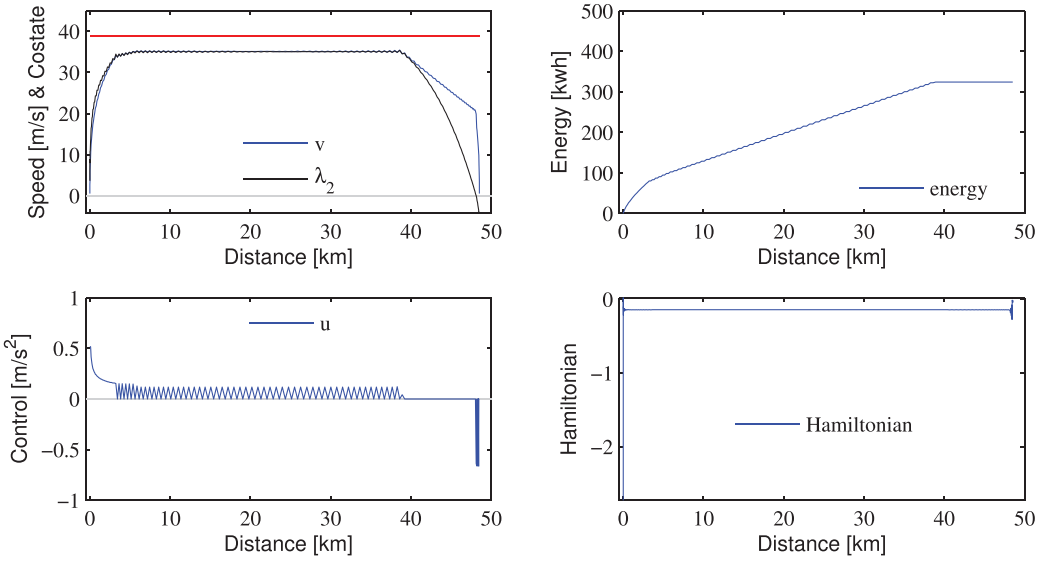

Fig. 5. Trajectories for the Intercity train with the energy-efficient driving strategy.

priate traction rather than finding the optimal switching points and switching between zero and maximum traction. A closer look at the control plot shows that the initial oscillating behaviour corresponding to the 'partial acceleration' part has a slightly higher amplitude in agreement with a higher maximum traction for lower speeds. So GPOPS has difficulties with the convergence of the control approximation at the beginning of the singular solution. Note that this oscillating control behaviour did not occur in the MTTC case. The essential difference is that in the MTTC case the singular solution occurred at an active state inequality path constraint, i.e., where the optimal cruising speed equals the maximum speed limit. This is a much easier case since the optimal cruising speed is then fixed.

The MTTC driving strategy of the SPR train consist of maximum acceleration, cruising at the speed limit and maximum braking (see Fig. 6), like the IC train. The total distance of cruising is shorter for the SPR train, due to the shorter total distance between the two stops. The results are in line with the necessary optimality conditions. The costate variable λ_2 behaves according to optimal control structure (27) although the control shows some unstable behaviour around the switching point from maximum acceleration to cruising. The Hamiltonian remains constant except for a small interval at the start. The same is observed in the Hamiltonian of the Sprinter EETC case. Experiments showed that slight changes to the scheduled running time T resulted in different deviations in the Hamiltonian in this interval suggesting that GPOPS has small numerical precision errors in computing the Hamiltonian in this region. In contrast, the states and costates are smooth while the control is also continuous. Table 2 indicates that the total amount of traction energy for this scenario is about 75 kWh. The results are computed within 2 s.

Finally, the EETC driving strategy for the SPR train with 15% running time supplements leads to a driving strategy without cruising, since the distance between the stops is too short to reach the optimal cruising speed, see Fig. 7. Therefore, only maximum acceleration, coasting and maximum braking are applied. Indeed, the EETC driving strategy for short distances with sufficient running time supplements consists of one acceleration to the optimal coasting speed after which the train coasts until a final braking regime to come to a standstill just in time. This is a gen-

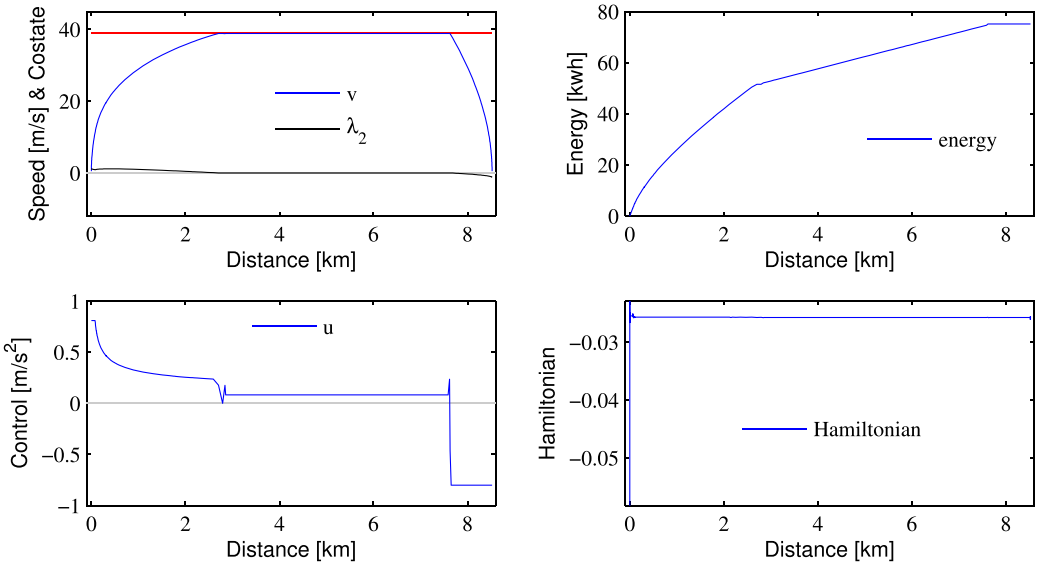

Fig. 6. Trajectories for the Sprinter train with the minimum-time driving strategy.

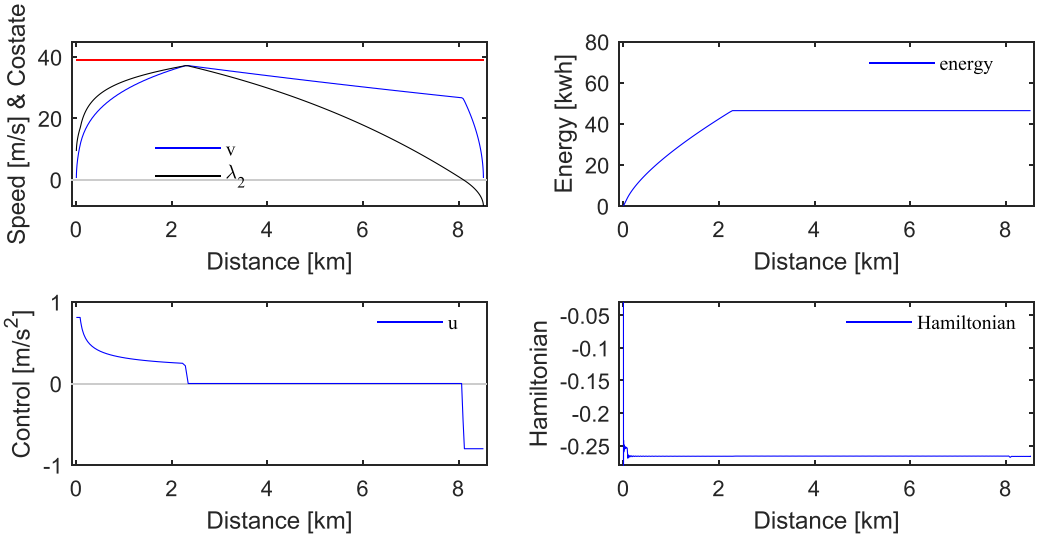

Fig. 7. Trajectories for the Sprinter train with the energy-efficient driving strategy.

eral energy-efficient driving strategy for urban trains (Scheepmaker et al., 2017). The costate λ_2 exceeds speed v if the train applies maximum acceleration, is between speed v and zero if the train is coasting, and below zero if the train applies maximum braking. These conditions are in line with the optimal control structure (19). Finally, the Hamiltonian remains constant. The total energy consumption of the EETC driving strategy of the SPR train is about 43% lower compared to the MTTC driving strategy (see Table 2). Finally, the results are generated in 15 s.

In summary, the results of the reference scenario are largely consistent with the necessary optimality conditions, except for the singular solution of the long distance IC train with an oscillating control approximating cruising by partial traction-coasting pairs, and some oscillation in the braking regime. The singular solution also starts a little bit too early where the speed still gradually increases until it oscillates around the optimal cruising speed. Alternating maximum acceleration and coasting is the optimal driving strategy for diesel-electric trains with finite discrete throttle settings (Howlett, 1996; 2000; Howlett & Pudney, 1995). For electric trains with continuous traction settings a constant traction according to the resistance forces is preferred from the viewpoint of driver workload and comfort, or an approximation with maximum traction-coast bang-bang control. The oscillating control computed by the pseudospectral method therefore must be understood as an approximation of the cruising speed only and not as the optimal control. This is discussed in more detail in Section 6. Finally, the MTTC problem is solved within seconds while the EETC problem takes more time. Clearly, the EETC problem is more involved as it needs to find the optimal cruising speed and coasting point.

5.3. Varying speed limits

This section considers the scenarios with varying speed limits causing discontinuous state path constraints (5). We only consider the EETC driving strategy with 15% running time supplements, and only the IC train since the results of the IC and SPR train are similar. In these scenarios we insert a speed restriction over the stretch between 25 km and 35 km. We gradually decrease the speed limit between these points from 140 to 120, 110 and 100 km/h. The resulting diagrams of the speed, costate, control and Hamiltonian

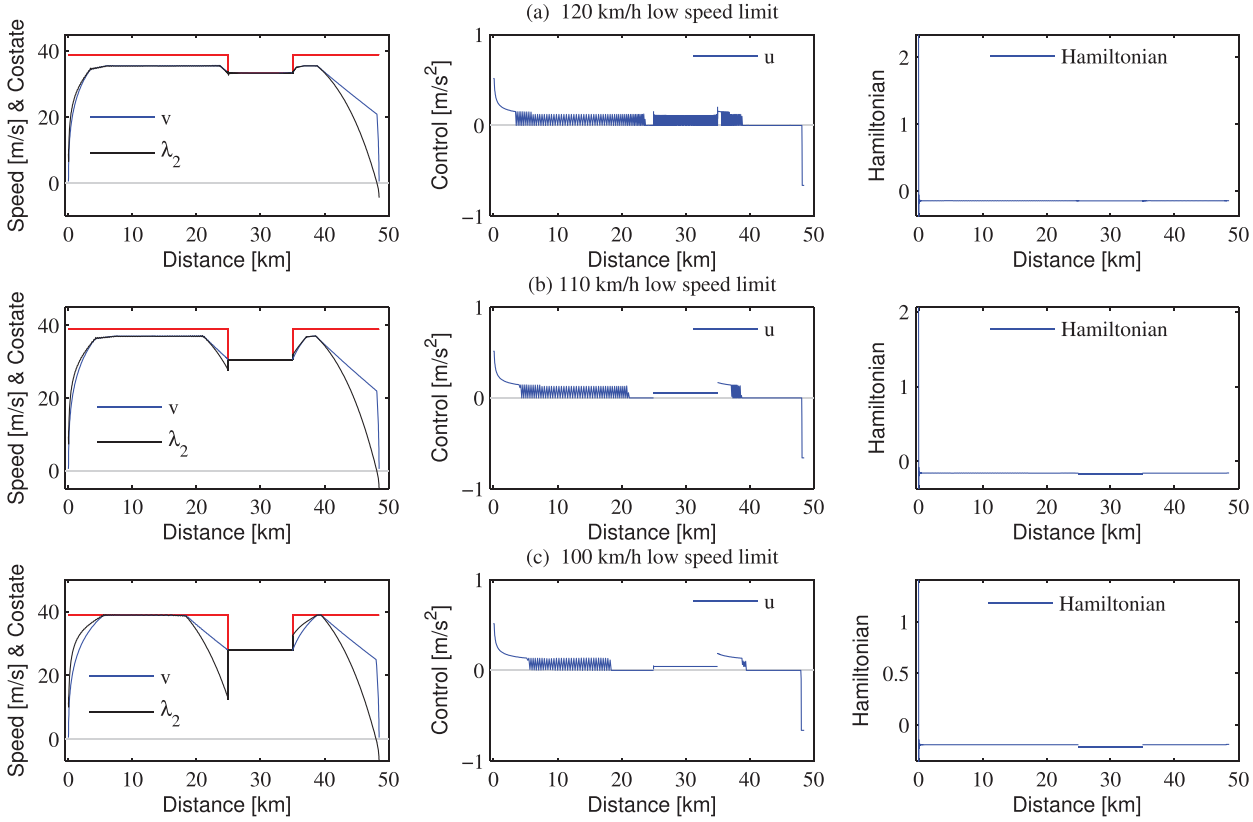

Fig. 8. Trajectories for the Intercity train with varying speed limits.

can be found in Fig. 8. The quantitative results are summarized in Table 2.

Fig. 8a shows the case where the speed limit between 25 km and 35 km is reduced to 120 km/h. The train applies maximum acceleration followed by cruising at an optimal speed below the speed limit, and then coasts before the speed restriction until the restricted speed is reached, which is then maintained until the end of the speed restriction after which the train accelerates maximally again to the optimal cruising speed, followed by coasting, and finally maximum braking. In the singular solutions, again an oscillating control is observed that approximates an optimal 'free' cruising speed below the speed limit, and also the singular solution at the restricted speed limit shows oscillating control to cruise below the restricted speed. Note that the cruising speeds before and after the speed restriction are the same, which is in line with the EETC theory that the optimal cruising speed is equal for each section (Howlett, 2016). The beginning of the singular solutions start again a bit too early with the oscillating speed creeping towards the optimal cruising speed, as we have observed in the IC EETC reference case. The costate λ_2 is consistent with the optimal control structure (19), however, it has now a jump at the begin and end of the speed restriction (this is more pronounced with the lower speed restrictions). Hence, the discontinuities in the pure state inequality path constraint for speed v, where the upper limit jumps from 140 km/h down to 120 km/h and back to 140 km/h, cause discontinuities to the costate λ_2. The value of the Hamiltonian remains constant over the complete trajectory. Table 2 shows that the speed restriction requires slightly more energy compared with the EETC reference case for the IC. The computation time increases to 130 s.

When the speed of the restriction is further reduced to 110 km/h and 100 km/h, the singular solution at the restricted speed limit consists of the optimal traction to cruise at the restricted speed, see Fig. 8b and 8c. The reduced speed leads to a longer running time over the speed restriction, which absorbs some of the running time supplement and less time is available before and after the speed restriction. As a result the optimal cruising speed goes up and coasting starts a bit earlier. In the 100 km/h scenario the cruising regime after the speed restriction takes longer before coasting can start since more time is wasted at the speed restriction. The cruising speeds before and after the speed restriction increase to the speed limit for lower restricted speeds, although the control plot of the 100 km/h case still shows oscillating control, so the computed cruising speed is here actually a bit below the speed limit and is approximated by oscillations. The jumps in the costate λ_2 at the speed limit changes are now clearly visible. Now also the Hamiltonians are discontinuous with a different value during the speed restriction, and equal ones for the first and third phase (see the values in Table 2). This is in line with the necessary optimality conditions described in Section 2.1. For big jumps in state constraints the costate approximation of λ_2 may no longer be accurate around the discontinuity (Darby et al., 2011a). For bigger speed drops (not shown here), λ_2 actually drops below zero although the train is coasting and not braking, so in these cases the KKT multiplier of the NLP solver no longer converges to the continuous costate around the speed jumps. Table 2 shows that more energy is required for decreasing speed restrictions. The computation time for these scenarios are about 117 and 90 s.

In summary, discontinuous state path constraints (varying speed limits) may cause discontinuities in the costate λ_2 and the Hamiltonian value at the positions of the state jumps (speed changes). This is in line with the necessary optimality conditions described in Section 2.1. However, the approximation λ_2 may no longer converge for big jumps in the state path constraints, and even become negative which could lead to erroneous conclusions

about the driving regime. Nevertheless, the state approximation behaved well around the nonsmooth state inequality constraints, and is not affected by the unstable control approximation at the jump points. Moreover, we observe that the oscillations of the singular control disappear when the speed restriction is further from the optimal cruising speed. Therefore, we hypothesize that the numerical issues of the singular control depend on the gap between the optimal cruising speed and the (local) speed limit for given collocation points. In Fig. 8a this gap is apparently not big enough for the speed restriction, while in Fig. 8b it is.

5.4. Varying gradients

This section considers varying gradients instead of a flat track, causing a discontinuous state equation (3). Since the results of IC and SPR train are similar, we only consider scenarios with the IC train, because these include the more complex regime of cruising at an optimal cruising speed below the speed limit in the EETC driving strategy. In these scenarios we consider a (steep) downhill slope or (steep) uphill slope between the distance from 25 km to 35 km. We will analyse them from less to more steep in the order -5%, -10%, 5% and 10%. The results can be found in Table 2 and Fig. 9. Note that in the figure the scenarios are ordered by increasing gradient.

Fig. 9b shows the results of a slight negative (downhill) gradient of -5% (0.005) between 25 km and 35 km. Because of the negative gradient, the train can apply cruising with less traction force to remain at the optimal cruising speed. The costate and control behave according to the optimal control structure, besides the slight increasing speed phenomenon at the beginning of the singular cruising, and the oscillating control approximating the cruising regime. The state and costate also show a small bump at the gradient jumps. The Hamiltonian is piecewise constant with a jump at the gradient (less negative), which is as expected from Eq. (8) of the Hamiltonian which includes the gradient. The energy savings are higher than the reference scenario, due to the downhill slope. Finally, the computation time is quite long with about 336 s so GPOPS has difficulties to find the optimal control that smoothes the effect of the gradient.

The scenario of the steep negative gradient of -10% (0.01) is shown in Fig. 9a. Here the section is so steep that the speed of the train increases, although it is coasting (zero traction). Therefore, the train starts to coast well before the downhill section to decrease speed before reaching the downhill slope and be able to stay below the speed limit on the negative slope where the speed gradually increases by the gravity forces. The coasting point has been optimized such that the train reaches the speed limit just at the bottom of the downhill section, and then continues with coasting until the optimal cruising speed is reached again. The optimal cruising speeds before and after the slope are increased to deal with the time loss during the long coasting regime before the slope. The costate λ_2 is in accordance with the optimal control structure (19), except around the gradient area. The approximated costate starts decreasing at the coasting point and keeps decreasing until a big discontinuity at the end of the slope where it jumps up to the speed limit to start cruising at the speed limit. But the costate even decreases below zero over a long distance, which would suggest maximum braking although the control shows coasting (zero traction). So the discontinuity of the dynamic equation causes an inaccurate approximation of the real optimal continuous costate λ_2. Nevertheless, the control and state show a correct behaviour. The Hamiltonian is piecewise constant with a less negative value on the slope, which is also as expected. Finally, a strange braking behaviour is observed at the end where the control shows a bang-bang control with maximum braking-coasting on a small distance although the costate is negative, so GPOPS has a difficulty here with determining the control. Under the optimal control the train should start braking a bit later and then maintain maximum braking until standstill, which means that the coasting should also start a bit later to arrive on time. The grid from the collocation points however limits the position of the switching points. From the state approximation, we can derive the energy consumption, which might be a bit too optimistic since coasting starts a bit too early. The steep downhill slope leads to more energy savings compared to the EETC reference scenario and the less steep slope scenario of the IC train. Table 2 indicates that the energy savings increased considerably to 51%. The computation time of this scenario is 136 s.

The effect of a slight positive (uphill) gradient of 5% (0.005) can be seen in Fig. 9c. The train maintains its optimal cruising speed over the uphill slope by applying more traction with oscillating control between two positive values to counter the higher resistance forces. The train already applies more traction slightly before the uphill slope, which causes a short peak in the speed. At the end of the section the speed drops shortly below the optimal cruising speed, which is caused by decreasing the traction again. The costate λ_2 behaves in accordance to the optimal control structure. The Hamiltonian is piecewise constant with discontinuities where the gradient changes. The Hamiltonian value is a bit lower over the uphill slope. The energy saving is less compared to the EETC reference scenario, due to the extra traction on the uphill slope (see Table 2). The computation time of this scenario is 135 s.

On the steep positive gradient of 10% (0.01), we see in Fig. 9d that the train has to apply maximum acceleration and still loses speed. Therefore, the train increases traction just before the slope and keeps at the maximum traction until after the slope when it has accelerated again to the optimal cruising speed. The costate λ_2 and the control are perfectly in line with the optimal control structure (19). The Hamiltonian is again piecewise constant with jumps at the changes of the gradient. By applying the EETC driving strategy the total energy is slightly smaller than the MTTC driving strategy of the IC reference scenario (see Table 2). The model results are computed within 163 s.

In summary, the experiments of the varying gradients (with jumps in the dynamic constraint) indicate that the EETC anticipates by adapting the control before the change of gradients. For a steep downhill slope where the gravity force exceeds the train resistance and the train accelerates despite zero traction, the optimal control starts coasting in advance such that the speed at the end of the slope is exactly the speed limit and then continuous coasting until the optimal cruising speed is reached again. For a steep uphill slope where the maximum traction is not sufficient to avoid slowing down, the train starts maximum acceleration before the slope and keeps this regime until after the slope when it is back to its optimal cruising speed. The Hamiltonian is piecewise constant with discontinuities at the gradient jumps. For steep negative gradients the costate λ_2 is not accurately approximated before the gradient jump, and also the control has difficulties to converge to the optimal control in the final braking regime. This may lead to an early coasting point that is corrected by additional coasting in the braking regime to arrive on time. Downhill slopes result in additional energy savings, while uphill slopes consume more energy. The varying gradients (discontinuous dynamic constraint) lead to higher computation times than the varying speed limits (discontinuous state path constraints).

5.5. Varying running time supplements

This subsection considers the effect of different running times T on the optimal EETC driving strategies by varying the amount of running time supplement, both for the IC and SPR train. The running time supplements are varied over 0%, 2%, 5%, 10%, 15%

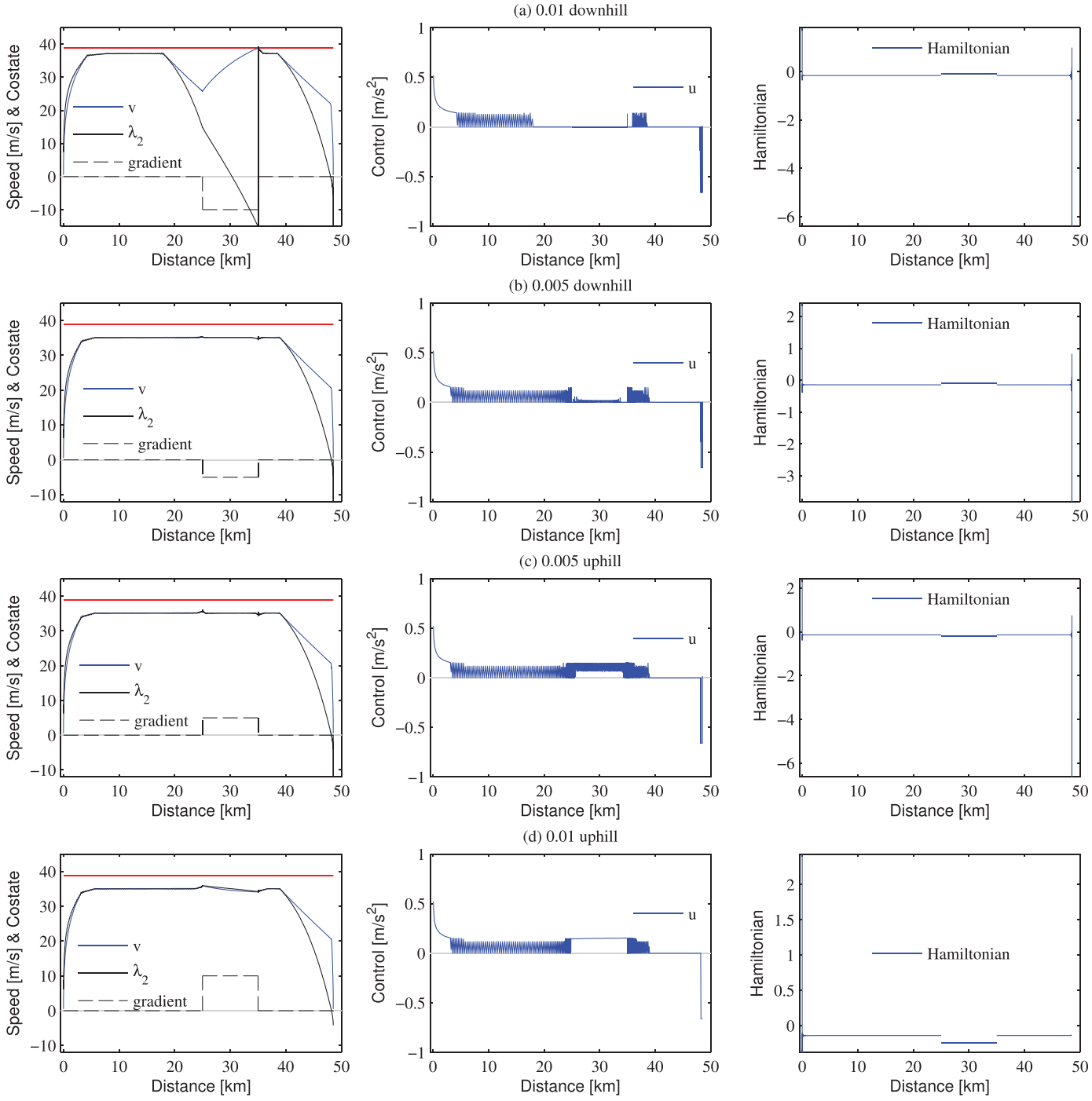

Fig. 9. Trajectories for the Intercity train with varying gradients (remark: gradients are indicative).

and 20%. The results are summarized in Table 2, and illustrated in Figs. 10 and 11 for the IC train and SPR train, respectively. The relation between energy consumption and total running time is also visualized in an energy-time diagram, which is shown on the right in the figures. The energy-time curve gives a Pareto frontier, which can be used to evaluate the solutions of a multiple-objective timetable problem (Domínguez, Fernández, Cucala, & Lukaszewicz, 2011).

The results for the IC train in Fig. 10 show two phenomena. First, for small supplements the optimal cruising speed equals the speed limit and a little supplement reduces the cruising regime quickly with an earlier switching point to coasting. Then from some sufficient supplement the optimal cruising speed starts decreasing below the speed limit and the switch to coasting starts later again. More supplement also leads to a later switch to braking at a lower speed. The energy savings increase up to 32% for 20% running time supplements compared to zero supplement. From the energy-time diagram, we see that the relative energy savings decrease for increasing supplement, i.e., adding supplement in the beginning leads to more energy savings than further increasing the supplement by the same amount later (convex curve). The approximations are consistent with the EETC optimal control structure, ex-

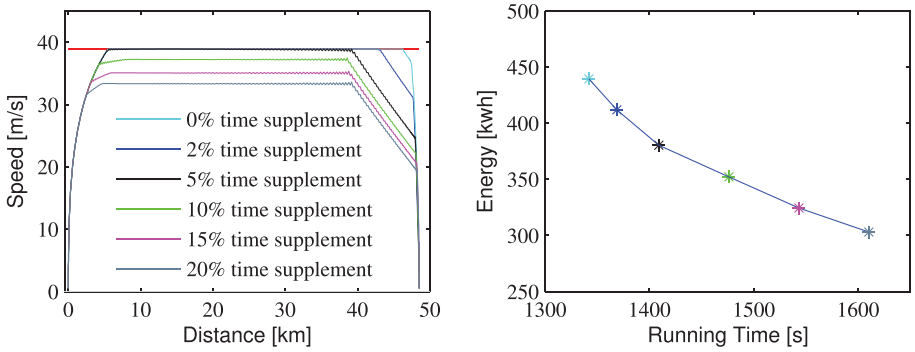

Fig. 10. Speed trajectories for the Intercity train with varying running times (left) and energy-time diagram (right).

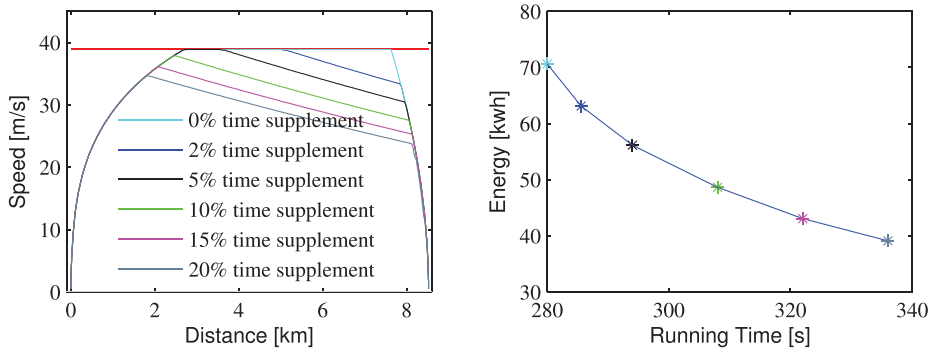

Fig. 11. Speed trajectories for the Sprinter train with varying running times (left) and energy-time diagram (right).

cept for the early start of the singular solution with the oscillating control similar to the EETC IC reference case with 15% supplement. The computation times vary between 10 s and 22 s.

The results for the SPR train in Fig. 11 show that with more running time supplement the coasting point is earlier, and with sufficient supplement the speed limit is no longer reached and the cruising regime is dismissed. In those cases, the optimal control switches from maximum acceleration to coasting, as was also seen in the SPR EETC reference case corresponding to the 15% supplement. The braking also starts later from a lower speed when more supplement is added. This behavior is in line with the optimal strategy for short distance trains consisting of maximum acceleration, coasting, and maximum braking (Scheepmaker et al., 2017). Again, the energy-time curve is convex with decreasing energy savings for increasing running time supplements. The energy savings increase up to 48% for 20% running time supplement compared to zero supplement. The computations are perfectly consistent with the EETC optimal control structure, similar to the EETC SPR reference case with 15% supplement. The computation time of the SPR train scenarios range between 15 s and 19 s.

In conclusion, increasing running times lead to extra energy savings for the EETC driving strategy with decreasing relative savings for increasing supplements as can be shown in an energy-time curve. The initial savings are very steep which reduces for higher supplements. When the optimal cruising speed is restricted by the speed limit for small supplements, coasting starts earlier for increasing supplements. Then for long distances and sufficient supplement (IC), the optimal cruising speed drops below the speed limit and coasting starts later again. For short distances and sufficient supplement (SPR), the optimal cruising speed is no longer reached and the optimal control switches from maximum acceleration to coasting directly. In all cases, more supplement leads to a later switch to braking from a lower speed. The pseudospectral method does not experience additional difficulties with respect to the EETC reference scenarios.

5.6. Real-life scenarios

The final scenarios are real-life cases of the IC and SPR trains. These cases include multiple speed limit changes as well as varying gradients caused by for example bridges, fly-overs and dive-unders. The applied running time supplement for the EETC scenarios is 10% of the computed minimum running time.

The results for the IC train between Ut and Ht are shown in Fig. 12 (MTTC) and Fig. 13 (EETC), and are summarized in Table 2. These figures indicate that the driving strategies behave according to the necessary optimality conditions. The minimum-time optimal control consists of maximum acceleration and cruising at the speed limit with partial traction or braking following the varying gradient resistances. Where a speed limit drops, the train uses maximum braking, cruising at the restricted speed limit, and maximum acceleration again after the speed limit increases. At the end, the train applies maximum braking in order to stop. The costate λ_2 also behaves according to the optimal control structure. Note that we did not include λ_2 in the real-life speed-distance profiles for clarity of the pictures. The Hamiltonian is piecewise constant with jumps where a speed limit or gradient changes. The costate $\lambda_1 = -1$. The computation time is 57 s.

The EETC driving strategy for the IC includes coasting regimes before speed limit reductions, which shortens the cruising and the braking regimes, and looks reasonable. The Hamiltonian is piecewise constant and the costate λ_1 is constant. The costate λ_2 also behaves according to the control structure. The EETC driving strategy with 10% running time supplements saves 42% in energy consumption compared to the MTTC driving strategy. The computation time of 203 s is quite high, due to the increased complexity of the varying speed limits and gradients.

The SPR real-life scenarios are shown in Fig. 14 (MTTC) and Fig. 15 (EETC), and are summarized in Table 2. The results correspond largely with the necessary optimality conditions. The MTTC driving strategy follows mainly the fastest running profile, with

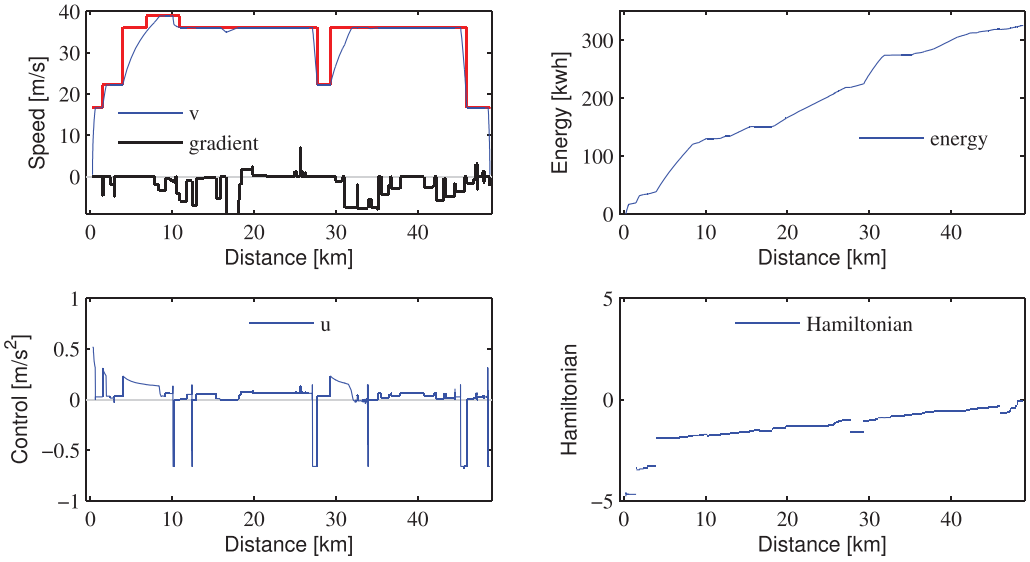

Fig. 12. Trajectories for the Intercity train with the minimum-time driving strategy on Ut-Ht (gradients indicative).

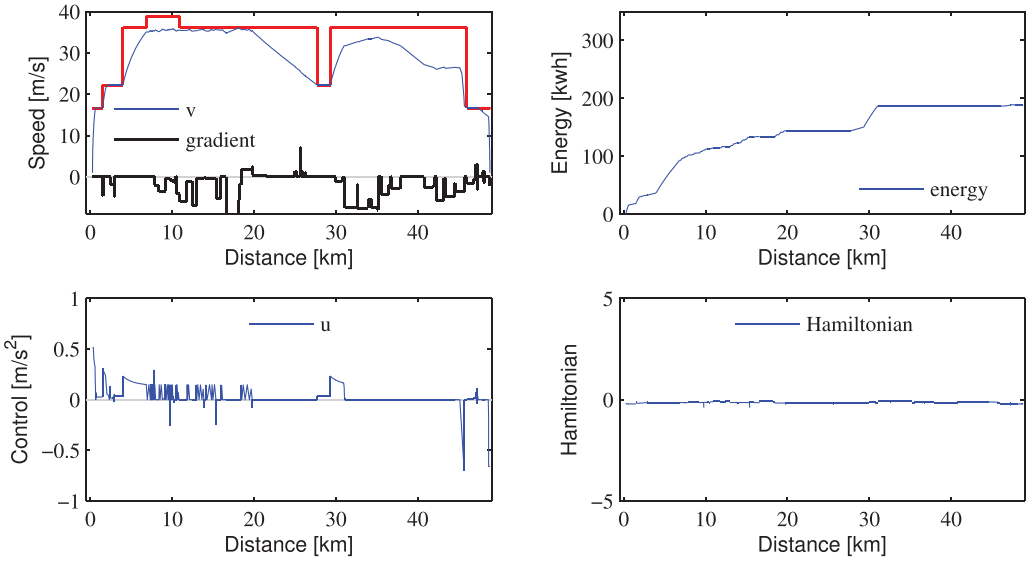

Fig. 13. Trajectories for the Intercity train with the energy-efficient driving strategy on Ut-Ht (gradients indicative).

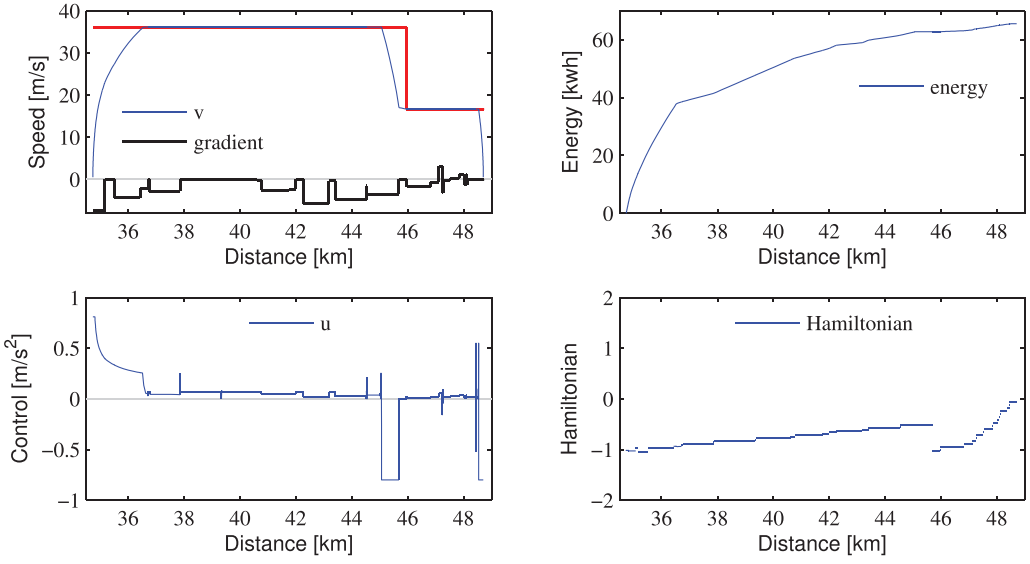

Fig. 14. Trajectories for the Sprinter train with the minimum-time driving strategy on Zbm-Ht (gradients indicative).

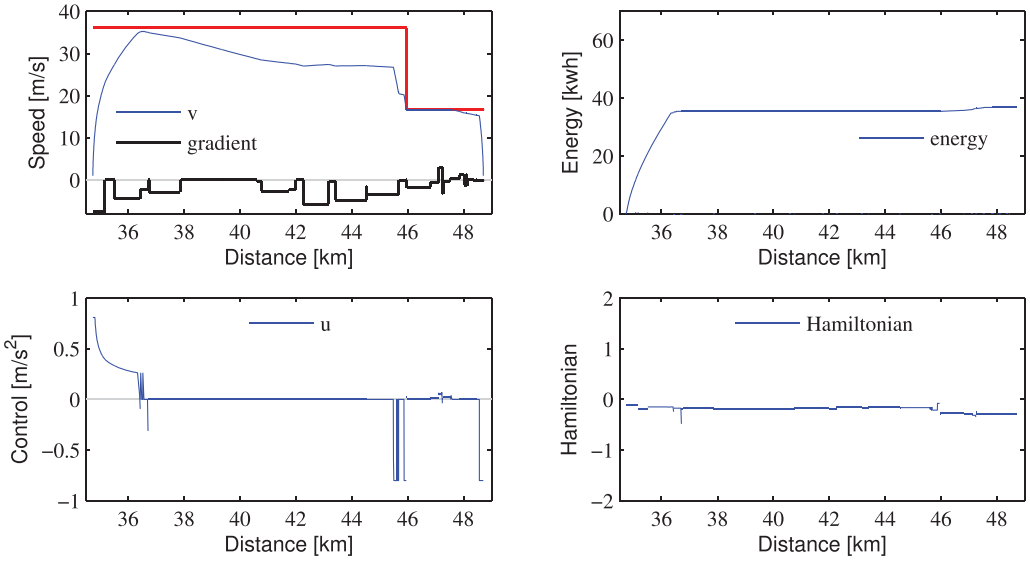

Fig. 15. Trajectories for the Sprinter train with the energy-efficient driving strategy on Zbm-Ht (gradients indicative).

maximum acceleration, cruising at the speed limit, maximum braking to decrease to the speed restriction, then an odd small coasting regime followed by cruising at the restricted speed, and maximum braking to standstill. The small coasting regime occurs at a flat track just before the speed restriction between two negative gradient sections. GPOPS thus has difficulties with the linkage conditions between these phases to brake correctly just before the speed restriction. The costate $\lambda_1 = -1$. The control in the cruising phases follows correctly the gradient resistance profile. The Hamiltonian is also piecewise constant following the discontinuities. The computation time is 36 s.

The SPR EETC driving strategy shows maximum acceleration followed by a short cruising regime and then a long coasting regime where the speed is influenced by the gradient profile. Then the control shows some unstable behaviour before the big speed restriction where the braking is interrupted with a coasting regime on the flat track part, similar to the MTTC. Note that the energy consumption is zero for both coasting and braking. Finally, the train cruises at the restricted speed, followed by coasting and maximum braking. The Hamiltonian is piecewise constant and the costate λ_1 is constant. The energy saving of the 10% running time supplement is computed as 44%. The computation time of the EETC model is about 63 s.

Based on the real-life case study it can be concluded that convergence issues may occur near big jumps in the state path constraints (speed restrictions) and dynamic equations (gradients), causing nonoptimal correcting behaviour of the control approximations (see Section 6). This coincides with the conclusions from Sections 5.3 and 5.4.

6. Solution to the singular oscillations

The previous section presented the results of the pseudospectral computations and compared them with the expectations from the PMP necessary optimality conditions. The computational results of the MTTC problems showed a good agreement with the PMP analysis and the computations were also very fast. In these cases, the optimal control structure is quite simple with running as fast as posssible where always an inequality constraint is active: either maximum traction force, maximum speed, or maximum braking. Also the EETC scenarios for the short-distance SPR train were in agreement with the PMP analysis without any major issues. In these cases, the maximum speed was either restricted by the speed limit or the train had to start coasting already before reaching the speed limit. The computational results for the EETC scenarios of the long-distance IC train, however, showed some disagreements with the PMP analysis, and therefore a larger range of scenarios were presented to analyse the impact of additional contraints to the computational results. The main difficulty of the IC EETC scenarios is the singular solution. Section 2.1 proved from the PMP conditions that the singular solution should correspond to a cruising regime in which an optimal cruising speed is maintained by partial traction. The pseudospectral method, however, computed an oscillating (discretized) control where the control jumps between two values over the successive collocation points, resulting also in oscillating behaviour of the state and costate trajectories. In this section, we analyse this behaviour in more detail and show that the oscillations should be understood as a discrete approximation of high-order global polynomials over the corresponding phases. Chen and Biegler (2016) added monotonic constraints to the control discretization to force low-order control profiles of the discrete approximation. A similar approach could also been done here, but this would mean an add-on to the pseudospectral method and, therefore, it would loose its appeal as a generic solver. Thus, we propose a new hybrid approach that combines the pseudospectral method and the PMP to compute accurate singular control solutions.

The pseudospectral method approximates the partial traction force corresponding to the optimal cruising speed by switching between two alternating values around the optimal traction over the successive collocation points. As can be observed from Figs. 5, 8 and 9, the lower value is mostly 0 m/s^2 and the upper value a (slightly varying) positive value. The only exception is the scenario of Fig. 9c, where the control oscillates between two positive values, including the upper bound on the control. In this case, the traction is more than half the control upper bound and can thus not be approximated by an oscillating control including zero traction. Note that the energy consumption is computed as a weighted sum over the (positive) control values at the collocation points. Hence, on a singular arc the oscillations approximate the cost by a weighted mean over successive collocation points. In other words, the discretized control values are not uniquely determined for the singular solution. When increasing the number of collocation points the magnitude of the jumps stay approximately the same, which makes sense when one of any two successive values is fixed to either the lower or upper bound. However, when

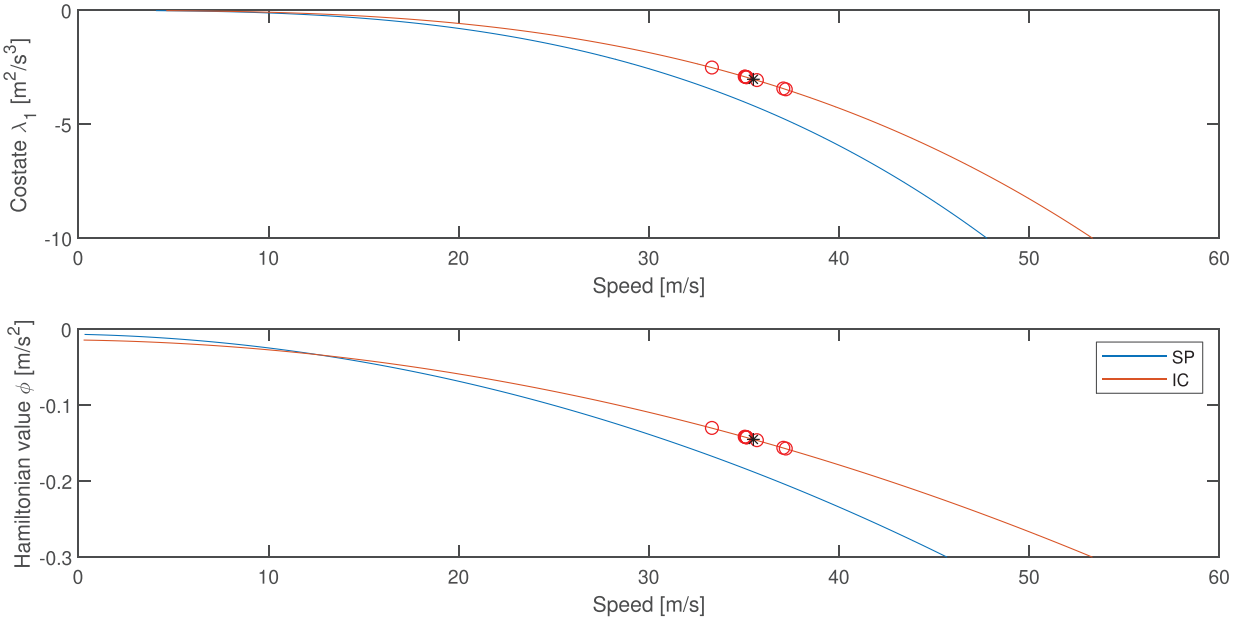

Fig. 16. Optimal cruising speeds as function of λ_1 and ϕ.

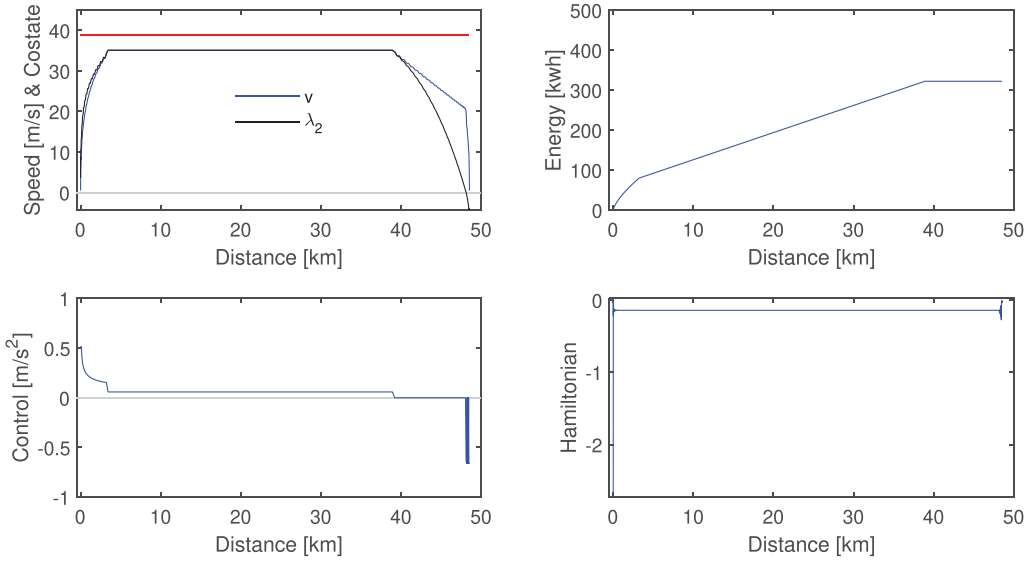

Fig. 17. Trajectories for the Intercity train with the energy-efficient driving strategy using the hybrid approach.

increasing the number of collocation points $N \to \infty$, it becomes clear that this is not a realistic traction control since the traction can not switch this fast between two values. The oscillating control should therefore be replaced by a low-order control representation corresponding to a feasible optimal traction with the same energy consumption as the discrete control approximation computed by the pseudospectral method. Likewise, the state and costate should be replaced by a low-order representation in the singular solution. In fact, from the PMP we know that the singular solution corresponds to a cruising regime with constant cruising speed and equal costate λ_2, so on the corresponding collocation points we must have $v(s) = \lambda_2(s) \equiv v_c$ with associated optimal control $u(s) = r(v_c) + g(s)$.

The optimal cruising speed can be computed from (17) and (18) depending implicitly on λ_1 and φ, respectively. Fig. 16 shows the analytical cruising speed as functions of λ_1 and φ for both the SPR and IC trains. Note that the PMP analysis does not provide a closed-form analytical expression for either λ_1 or φ, and in fact, the implicit equations are underdetermined, i.e., there are two nonlinear equations in three unknowns λ_1, φ and v_c. However, when one of the unknowns is given, the other two can also be determined. The pseudospectral method computes both λ_1 and φ, and thus we can also compute the corresponding optimal cruising speed using the implicit equations (17) and (18). In Fig. 16, the circles correspond to the resulting cruising speed v_c for the computed λ_1 and φ for the eight IC EETC scenarios where the cruising speed stayed below the speed limit, while the black stars correspond to the real-life IC EETC scenario. For each scenario, the values of λ_1 and φ computed by the pseudospectral method show a good fit to the analytical optimal cruising speed curves. Note also that the values of λ_1 and φ in Table 2 for the SPR EETC scenarios all correspond to optimal cruising speeds above the speed limit so that they indeed were never reached. Fig. 16 indicates that the optimal cruising speed is a decreasing concave function of either λ_1 or φ, with more negative values of λ_1 and φ corresponding to higher cruising speeds. The sensitivity of the cruising speed increases for

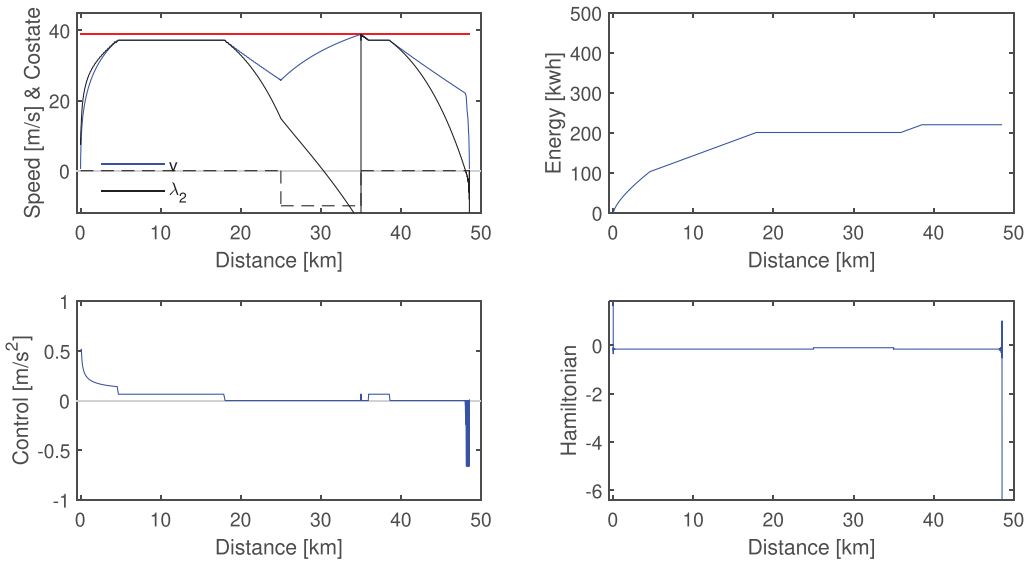

Fig. 18. Trajectories for the Intercity train with the energy-efficient driving strategy on a partial steep downhill track using the hybrid approach.

small absolute values of λ_1 and φ, which however correspond to extremely high running time supplements. For example, the excessive running time supplement of 20% in the IC and SPR case studies correspond to a cruising speed of 33.32 m/s and 34.69 m/s, respectively, which are far away from the sensitive area. Since λ_1 is a constant, it is recommended to use (17) for computing the optimal cruising speed based on the computed value of λ_1 by the pseudospectral method. The computed value of φ can be used for checking the accuracy together with the gradient profile.

The singular solution can thus be obtained by a combination of the PMP and the pseudospectral method using the following algorithm.

1. Apply the multiple-phase pseudospectral method to find the optimal value λ_1, and the discretized solution $\{(t(s_i), v(s_i), \lambda_2(s_i), u(s_i)) \mid i \in \{0, \ldots, N\}\}$ with $N = \sum_{r=1}^{R} N_r$.
2. Find the optimal cruising speed v_c by solving the implicit equation (17) for given λ_1 using any root-finding algorithm (e.g. the MATLAB function fzero).
3. Detect singular solutions associated to discretization points s_i with $|v(s_i) - \lambda_2(s_i)| < \varepsilon$ for some small ε (e.g. $\varepsilon = 0.01$).
4. On the singular points set $v(s_i) = \min(v_c, v_{\max}(s_i))$, $\lambda_2(s_i) = \min(v_c, v_{\max}(s_i))$, $u(s_i) = r(\min(v_c, v_{\max}(s_i))) + g(s_i)$.

Fig. 17 shows the trajectories of the reference IC EETC scenario using the hybrid approach, cf. Fig. 5 for the original pseudospectral solution. Note that the solution satisfies the PMP optimality conditions as can be observed from the plots. For the IC EETC reference scenario the cruising speed computed by the pseudospectral method is approximately 35.12 m/s, i.e., the discrete control approximation fluctuates around this value, while the analytical optimal cruising speed for the computed $\lambda_1 = -2.9256$ is $v_c = 35.10$ m/s^2. For this speed $\varphi = -0.1423$ using (18), while the pseudospectral method computed an approximation of -0.1422 (with some slight fluctuations around the fourth decimal). This $\lambda_1 = -2.9256$ corresponds to 15% running time supplement over the minimum running time. In Table 2, we can see that decreasing the supplement leads to more negative λ_1 (larger absolute value) corresponding to a higher cruising speed. This makes sense as less time is available to reach the destination and so the train must run faster. When reducing the speed restriction, we also observe an increasing $|\lambda_1|$ indicating that less supplement is available and the optimal cruising speed thus increases where possible. A speed restriction of 110 km/h reduces the running time supplement to about 10% and a lower speed limit of 100 km/h further reduces the supplement to less than 5%. The steep negative gradient of -10% also leads to a more negative $\lambda_1 = -3.4671$. In this case, the cruising speed increases to compensate for the long coasting before, over and after the downhill slope. Fig. 18 shows the corrected trajectories for this scenario using the hybrid method, cf. Fig. 9a for the original pseudospectral solution. For the real-life IC EETC scenario $\lambda_1 = -3.0222$, which corresponds to an optimal cruising speed of $v_c = 35.50$ m/s (127.8 km/h), and $\varphi = -0.1450$ on flat track, see the black stars in Fig. 16.

Replacing the pseudospectral solution of the trajectories in the cruising regime by the analytical solution from the implict equation(17) for the computed λ_1 also solves the 'partial acceleration' issue at the beginning of the cruising regime. The exact switching point from the acceleration regime to the cruising regime can be computed as the point where the maximum acceleration curve reaches the analytical cruising speed. The braking regime also shows some oscillations between zero and maximum braking. These inaccuracies can be explained from the discretization. The distribution of LGR collocation points over the total distance is fixed for each N. The optimal switching point from maximum acceleration to cruising (or coasting) is usually not located exactly at a collocation point, which thus generates an approximation error. Likewise for the other switching points from cruising to coasting and from coasting to braking. Since the total running time is fixed and included as a hard condition, the pseudospectral method compensates the imprecision of the discretization by a nonoptimal partial traction regime over a number of collocation points before reaching the real optimal cruising speed regime. And likewise, a partial braking regime at the end compensates for the imprecision of the switching point from coasting to braking. This inaccuracy can be solved by increasing the number of collocation points but this increases the computation time. An alternative is to add extra phases around the switching points in a second iteration of the pseudospectral method, which thus gives a finer collocation grid to better determine the optimal switching points. For a maximum braking regime also another postprocessing approach can be used based on the PMP. The exact switching point to the maximum braking regime can be computed as the point where the coasting speed curve reaches the maximum braking curve (calculated backwards), and thus removing any partial braking within the braking phase. Any accuracy error due to the collocation grid for the coast-

ing and braking points can be corrected by optimizing the switching point from cruising to coasting such that the total running time equals the scheduled running time T. The main challenge of finding the optimal cruising speed v_c is however solved by the pseudospectral method via the computed value of the costate λ_1.

7. Conclusions

In this paper we applied the pseudospectral method to two optimal train control problems with speed and time as state functions of distance. We proposed a framework connecting the pseudospectral method with the PMP and showed that the pseudospectral solutions were in general consistent with the necessary optimality conditions but also identified some convergence issues. The main advantage of using the pseudospectral method is that no *a priori* knowledge is needed about the optimal control structure. We first applied the PMP to both the EETC and MTTC problem, providing the necessary optimality conditions. The necessary conditions provide information about the optimal control structure in terms of speed and the costate λ_2 associated to speed, as well as characteristics of the costate λ_1 associated to time and the Hamiltonian value. These were used to validate and correct the solutions computed by the pseudospectral method.

We used the MATLAB toolbox GPOPS where the Radau pseudospectral method with multiple-phases is implemented to solve the optimal control problems. Within each phase the gradients and speed limits remain constant. We applied our framework to various scenarios for both Intercity and Sprinter trains. The pseudospectral method works well for the MTTC problem, and also for the EETC problem when either (1) the cruising regime is absent due to a short distance or sufficient time supplement in which cases the optimal control switches from acceleration directly to coasting before reaching the optimal cruising speed; or (2) the speed limit becomes active before reaching the theoretical optimal cruising speed. However, the pseudospectral method has numerical issues with singular cruising solutions when the optimal cruising speed is below the speed limit or close above the speed limit depending on the number of collocation points. These cruising regimes are approximated by oscillations over the successive collocation points with alternating partial-zero traction or partial-full traction regimes and may start a bit too early before reaching the cruising speed. Also the final maximum baking regime may start with partial braking before full braking and likewise the switch from cruising to coasting may suffer from a small section of inaccurate partial traction. In these cases, the optimal switching points are located between the collocation points and therefore the offset of the optimal switching point needs to be corrected by a small regime of some appropriate non-optimal control values. Increasing the number of collocation points can correct this at the cost of higher computation time. Furthermore, discontinuities in the dynamic equation (gradients) or state path constraint (speed limits) lead to a discontinuous costate λ_2 with jumps at the points of discontinuities. For big negative jumps in the dynamic equation or state path constraint the costate λ_2 may not always be approximated correctly and the optimal control structure is violated before the discontinuities, while also the control may have difficulties to converge. So care has to be taken in using the costate and control around serious discontinuities.

The approximate oscillating singular solutions can be corrected in a postprocessing step using an implicit nonlinear equation from Pontryagin's Maximum Principle. This implicit equation is underdetermined but the unknown λ_1 is computed as part of the pseudospectral solution. With this value of λ_1, the optimal cruising speed can be computed numerically from the implicit nonlinear equation using a root-finding algorithm. Possibly corrected by an active speed limit, the constant cruising speed can replace the numerical approximations of the state v and costate λ_2 at the collocation points in the singular solution directly. The optimal control can then also be derived analytically from the cruising speed using the dynamic equation in equilibrium.

In the optimal control structure, the optimal cruising control would have to adjust at each gradient change which may result in some time delay unless the gradient profile is perfectly known and the control acts perfectly in line with it. In practice, the driver or an automatic speed control system will adaptively control the traction to hold the cruising speed and thus respond to changes in resistance, such as changes in gradients, curves and wind speed. The optimal control structure (19) is thus an open-loop control that provides a feedforward reference train trajectory. More important are the associated speed trajectory and driving regimes that a driver or speed control system can use as target. So rather than implementing the theoretical control (19), the actual maximum traction and maximum (service) braking should be used, as well as the proper traction control that maintains the cruising speed and in essence equals the actual resistance forces.

We conclude that the Radau pseudospectral method provides a direct general and flexible means to solve optimal train control problems, especially when the optimal control structure is not known beforehand (more complex situations with many constraints). Combining the pseudospectral solution with key equations from the necessary optimality conditions of Pontryagin's Maximum Principle improves the solution inaccuracy due to the discretization. Our paper also advocates solid verification and validation. A lot of literature on optimal train control applies heuristics without proper verification or validation of the resulting solutions. A comparison to the PMP necessary optimality conditions as done in this paper is highly recommendable. However, this is mathematically demanding and is also less powerful when the costates are not computed. An alternative approach is to use the pseudospectral method with available software like GPOPS to compare the solutions and improve algorithms. This paper can then be used as a guideline where to be careful in the interpretation of the results of the pseudospectral method.

Acknowledgments

The authors like to thank the Netherlands Railways (NS) and ProRail for making this research possible. Moreover, the authors also like to thank the anonymous reviewers for their useful comments.

References

Albrecht, A. R., Howlett, P. G., Pudney, P. J., Vu, X., & Zhou, P. (2016a). The key principles of optimal train control–part 1: Formulation of the model, strategies of optimal type, evolutionary lines, location of optimal switching points. *Transportation Research Part B: Methodological, 94*, 482–508. https://doi.org/10.1016/j.trb.2015.07.023.

Albrecht, A. R., Howlett, P. G., Pudney, P. J., Vu, X., & Zhou, P. (2016b). The key principles of optimal train control–part 2: Existence of an optimal strategy, the local energy minimization principle, uniqueness, computational techniques. *Transportation Research Part B: Methodological, 94*, 509–538. https://doi.org/10.1016/j.trb.2015.07.024.

Benson, D. A., Huntington, G. T., Thorvaldsen, T. P., & Rao, A. V. (2006). Direct trajectory optimization and costate estimation via an orthogonal collocation method. *Journal of Guidance, Control, and Dynamics, 29*(6), 1435–1440.

Bertsekas, D. P. (1999). *Nonlinear programming*. Belmont, MA, USA: Athena Scientific.

Betts, J. T. (2010). Practical methods for optimal control and estimation using nonlinear programming. *Advances in Design and Control*. Philadelphia, PA, USA: SIAM. https://doi.org/10.1137/1.9780898718577.

Brünger, O., & Dahlhaus, E. (2014). Running time estimation. In I. A. Hansen, & J. Pachl (Eds.), *Railway timetabling & operations* (pp. 65–89). Hamburg, Germany: Eurailpress.

Bryson, A. E., & Ho, Y.-C. (1975). *Applied optimal control*. New York: Taylor & Francis.

Chen, W., & Biegler, L. T. (2016). Nested direct transcription optimization for singular optimal control problems. *AIChE Journal, 62*(10), 3611–3627. https://doi.org/10.1002/aic.15272.

Chevrier, R., Pellegrini, P., & Rodriguez, J. (2013). Energy saving in railway timetabling: A bi-objective evolutionary approach for computing alternative running times. *Transportation Research Part C: Emerging Technologies, 37*, 20–41. https://doi.org/10.1016/j.trc.2013.09.007.

Darby, C. L., Garg, D., & Rao, A. V. (2011a). Costate estimation using multiple-interval pseudospectral methods. *Journal of Spacecraft and Rockets, 48*(5), 856–866.

Darby, C. L., Hager, W. W., & Rao, A. V. (2011b). An *hp*-adaptive pseudospectral method for solving optimal control problems. *Optimal Control Applications and Methods, 32*(4), 476–502.

Davis, W. J. (1926). The tractive resistance of electric locomotives and cars. *General Electric Review, 29*, 685–708.

Domínguez, M., Fernández, A., Cucala, A. P., & Lukaszewicz, P. (2011). Optimal design of metro automatic train operation speed profiles for reducing energy consumption. *Proceedings of the Institution of Mechanical Engineers, Part F: Journal of Rail and Rapid Transit, 225*(5), 463–474.

Elnagar, G., Kazemi, M. A., & Razzaghi, M. (1995). The pseudospectral Legendre method for discretizing optimal control problems. *IEEE Transactions on Automatic Control, 40*(10), 1793–1796. https://doi.org/10.1109/9.467672.

Fahroo, F., & Ross, I. M. (2001). Costate estimation by a Legendre pseudospectral method. *Journal of Guidance, Control, and Dynamics, 24*(2), 270–277. https://doi.org/10.2514/2.4709.

Garg, D. (2011). *Advances in global pseudospectral methods for optimal control*. Gainesville, FL, USA: University of Florida PhD thesis.

Garg, D., Patterson, M. A., Darby, C. L., Francolin, C., Huntington, G. T., Hager, W. W., & Rao, A. V. (2011). Direct trajectory optimization and costate estimation of finite-horizon and infinite-horizon optimal control problems using a Radau pseudospectral method. *Computational Optimization and Applications, 49*(2), 335–358. https://doi.org/10.1007/s10589-009-9291-0.

Garg, D., Patterson, M. A., Hager, W. W., Rao, A. V., Benson, D. A., & Huntington, G. T. (2009). An overview of three pseudospectral methods for the numerical solution of optimal control problems. *Advances in the Astronautical Sciences, 135*(1), 475–487.

Garg, D., Patterson, M. A., Hager, W. W., Rao, A. V., Benson, D. A., & Huntington, G. T. (2010). A unified framework for the numerical solution of optimal control problems using pseudospectral methods. *Automatica, 46*(11), 1851–1883.

Gong, Q., Fahroo, F., & Ross, I. (2008a). Spectral algorithm for pseudospectral methods in optimal control. *Journal of Guidance, Control, and Dynamics, 31*(3), 460–471. https://doi.org/10.2514/1.32908.

Gong, Q., Ross, I. M., Kang, W., & Fahroo, F. (2008b). Connections between the covector mapping theorem and convergence of pseudospectral methods for optimal control. *Computational Optimization and Applications, 41*(3), 307–335. https://doi.org/10.1007/s10589-007-9102-4.

Haahr, J. T., Pisinger, D., & Sabbaghian, M. (2017). A dynamic programming approach for optimizing train speed profiles with speed restrictions and passage points. *Transportation Research Part B: Methodological, 99*, 167–182. https://doi.org/10.1016/j.trb.2016.12.016.

Howlett, P. G. (1996). Optimal strategies for the control of a train. *Automatica, 32*(4), 519–532. https://doi.org/10.1016/0005-1098(95)00184-0.

Howlett, P. G. (2000). The optimal control of a train. *Annals of Operations Research, 98*(1), 65–87. https://doi.org/10.1023/a:1019235819716.

Howlett, P. G. (2016). A new look at the rate of change of energy consumption with respect to journey time on an optimal train journey. *Transportation Research Part B: Methodological, 94*, 387–408. https://doi.org/10.1016/j.trb.2016.10.004.

Howlett, P. G., & Pudney, P. J. (1995). *Energy-efficient train control*. London, UK: Springer.

Howlett, P. G., Pudney, P. J., & Vu, X. (2009). Local energy minimization in optimal train control. *Automatica, 45*(11), 2692–2698. https://doi.org/10.1016/j.automatica.2009.07.028.

Huntington, G. T. (2007). *Advancement and analysis of a Gauss pseudospectral transcription for optimal control problems*. Cambridge, MA, USA: Massachusetts Institute of Technology (MIT) PhD thesis.

Khmelnitsky, E. (2000). On an optimal control problem of train operation. *IEEE Transactions on Automatic Control, 45*(7), 1257–1266. https://doi.org/10.1109/9.867018.

Lewis, F. L., Vrabie, D. L., & Syrmos, V. L. (2012). *Optimal control*. Hoboken, NJ, USA: Wiley.

Liu, R. R., & Golovitcher, I. M. (2003). Energy-efficient operation of rail vehicles. *Transportation Research Part A: Policy and Practice, 37*(10), 917–932. https://doi.org/10.1016/j.tra.2003.07.001.

NS (2017). *TreinPlein* https://treinplein.net.

Patterson, M. A., Hager, W. W., & Rao, A. V. (2015). A *ph* mesh refinement method for optimal control. *Optimal Control Applications and Methods, 36*(4), 398–421. https://doi.org/10.1002/oca.2114.

Patterson, M. A., & Rao, A. V. (2014). GPOPS-II: A MATLAB software for solving multiple-phase optimal control problems using *hp*-adaptive gaussian quadrature collocation methods and sparse nonlinear programming. *ACM Transactions on Mathematical Software, 41*(1), 1:1–1:37. https://doi.org/10.1145/2558904.

Pontryagin, L., Boltyanskii, V. G., Gamkrelidze, R. V., & Mishchenko, E. F. (1962). *The mathematical theory of optimal processes*. Hoboken, NY, USA: Wiley.

Rao, A. V. (2014). Trajectory optimization: A survey. In H. Waschl, I. Kolmanovsky, M. Steinbuch, & L. del Re (Eds.), *Optimization and optimal control in automotive systems* (pp. 3–21). Cham, Germany: Springer. https://doi.org/10.1007/978-3-319-05371-4_1.

Rao, A. V., Benson, D. A., Darby, C., Patterson, M. A., Francolin, C., Sanders, I., & Huntington, G. T. (2010). Algorithm 902: GPOPS, a MATLAB software for solving multiple-phase optimal control problems using the Gauss pseudospectral method. *ACM Transactions on Mathematical Software (TOMS), 37*(2), 22:1–22:39.

Ross, I. M. (2005). A historical introduction to the covector mapping principle. In *Advances in the astronautical sciences: 123* (pp. 1257–1278). San Diego, CA, USA: Univelt Inc..

Ross, I. M. (2015). *A Primer on Pontryagin's Principle in Optimal Control*. San Fransisco, CA, USA: Collegiate Publishers.

Ross, I. M., & Fahroo, F. (2004). Pseudospectral knotting methods for solving optimal control problems. *Journal of Guidance, Control, and Dynamics, 27*(3), 397–405. https://doi.org/10.2514/1.3426.

Ross, I. M., & Karpenko, M. (2012). A review of pseudospectral optimal control: from theory to flight. *Annual Reviews in Control, 36*(2), 182–197.

Scheepmaker, G. M., & Goverde, R. M. P. (2015). The interplay between energy-efficient train control and scheduled running time supplements. *Journal of Rail Transport Planning & Management, 5*(4), 225–239. https://doi.org/10.1016/j.jrtpm.2015.10.003.

Scheepmaker, G. M., & Goverde, R. M. P. (2016). Energy-efficient train control including regenerative braking with catenary efficiency. In *IEEE international conference on intelligent rail transportation (ICIRT)* (pp. 116–122).

Scheepmaker, G. M., Goverde, R. M. P., & Kroon, L. G. (2017). Review of energy-efficient train control and timetabling. *European Journal of Operational Research, 257*(2), 355–376. https://doi.org/10.1016/j.ejor.2016.09.044.

Sicre, C., Cucala, A. P., & Fernández-Cardador, A. (2014). Real time regulation of efficient driving of high speed trains based on a genetic algorithm and a fuzzy model of manual driving. *Engineering Applications of Artificial Intelligence, 29*, 79–92. https://doi.org/10.1016/j.engappai.2013.07.015.

Wang, P., & Goverde, R. M. P. (2016a). Multiple-phase train trajectory optimization with signalling and operational constraints. *Transportation Research Part C: Emerging Technologies, 69*, 255–275. https://doi.org/10.1016/j.trc.2016.06.008.

Wang, P., & Goverde, R. M. P. (2016b). Two-train trajectory optimization with a green wave policy. *Transportation Research Record, 2546*, 112–120.

Wang, P., & Goverde, R. M. P. (2017). Multi-train trajectory optimization for energy efficiency and delay recovery on single-track railway lines. *Transportation Research Part B: Methodological, 105*, 340–361.

Wang, P., & Goverde, R. M. P. (2019). Multi-train trajectory optimization for energy–efficient timetabling. *European Journal of Operational Research, 272*(2), 621–635.

Wang, Y., DeSchutter, B., Van den Boom, T. J. J., & Ning, B. (2013). Optimal trajectory planning for trains – a pseudospectral method and a mixed integer programming approach. *Transportation Research Part C: Emerging Technologies, 29*, 97–114.

Wang, Y., DeSchutter, B., Van den Boom, T. J. J., & Ning, B. (2014). Optimal trajectory planning for trains under fixed and moving signaling systems using mixed integer linear programming. *Control Engineering Practice, 22*, 44–56.

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. https://doi.org/10.1016/j.trb.2016.08.002.

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. https://doi.org/10.1016/j.trc.2017.06.011.

Zhong, W., Lin, Q., Loxton, R., & Teo, K. L. (2019). Optimal train control via switched system dynamic optimization. *Optimization Methods and Software*. https://doi.org/10.1080/10556788.2019.1604704.