CrossMark

The key principles of optimal train control—Part 1: Formulation of the model, strategies of optimal type, evolutionary lines, location of optimal switching points

Amie Albrecht*, Phil Howlett, Peter Pudney, Xuan Vu, Peng Zhou

Scheduling and Control Group, Centre for Industrial and Applied Mathematics, Barbara Hardy Institute, University of South Australia, Mawson Lakes 5095, Australia

ARTICLE INFO

ABSTRACT

We discuss the problem of finding an energy-efficient driving strategy for a train journey on an undulating track with steep grades subject to a maximum prescribed journey time. We review the state-of-the-art and establish the key principles of optimal train control for a general model with continuous control. The model with discrete control is not considered. We assume only that the tractive and braking control forces are bounded by non-increasing speed-dependent magnitude constraints and that the rate of energy dissipation from frictional resistance is given by a non-negative strictly convex function of speed. Partial cost recovery from regenerative braking is allowed. The cost of the strategy is the mechanical energy required to drive the train. Minimising the mechanical energy is an effective way of reducing the fuel or electrical energy used by the traction system. The paper is presented in two parts. In Part 1 we discuss formulation of the model, determine the characteristic optimal control modes, study allowable control transitions, establish the existence of optimal switching points and consider optimal strategies with speed limits. We find algebraic formulae for the adjoint variables in terms of speed on track with piecewise-constant gradient and draw phase plots of the associated optimal evolutionary lines for the state and adjoint variables. In Part 2 we will establish important integral forms of the necessary conditions for optimal switching, find general bounds on the positions of the optimal switching points, justify the local energy minimization principle and show how these ideas are used to calculate optimal switching points. We will prove that an optimal strategy always exists and use a perturbation analysis to show the strategy is unique. Finally we will discuss computational techniques in realistic examples with steep gradients and describe typical optimal strategies for a complete journey.

1. Introduction

The modern theory of optimal train control has been developed during the years 1992–2014 by the Scheduling and Control Group (SCG) at the University of South Australia in a collection of papers—listed in chronological order—by Cheng and Howlett (1992), Howlett and Cheng (1993), Howlett et al. (1994), Howlett (1996, 2000), Howlett and Jiaxing (1997), Howlett and Leizarowitz (2001), Howlett et al. (2009) and Albrecht et al. (2013b), and more or less concurrently in elegant papers by

* Corresponding author. Tel.: +61 883023754.
E-mail addresses: amie.albrecht@unisa.edu.au (A. Albrecht), phil.howlett@unisa.edu.au (P. Howlett), peter.pudney@unisa.edu.au (P. Pudney), xuan.vu@unisa.edu.au (X. Vu), peng.zhou@unisa.edu.au (P. Zhou).
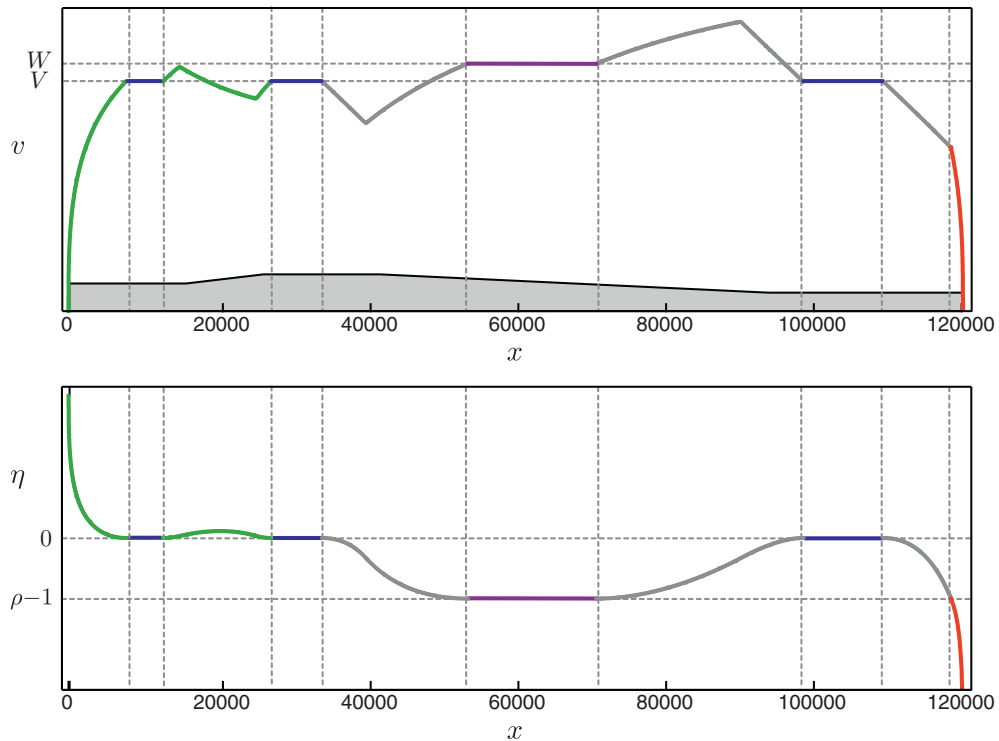
Fig. 1. Optimal speed profile $v = v(x)$ where x denotes position (top graph) with (i) initial phase of Maximum Power; (ii) extended HoldP phase with $V = 25.00$ showing inserted phases of Maximum Power, Coast and HoldR with $W = 26.93$; (iii) semi-final Coast phase; and (iv) final Maximum Brake phase. The corresponding modified adjoint profile $\eta = \eta(x)$ (bottom graph) shows the critical values $\eta = 0$ and $\eta = \rho - 1$ where $\rho = 0.8$ is the proportion of mechanical energy recovered during regenerative braking. The vertical dotted lines indicate points where the control changes. The position x is measured in metres (m), the speed v is measured in metres per second (m s^{-1}) and the modified adjoint variable η is dimensionless.

Khmelnitsky (2000) and by Liu and Golovitcher (2003). The book on energy-efficient train control by Howlett and Pudney (1995) is also frequently cited. Additional references of lesser significance can be found in many of these works. Some SCG papers consider the case of discrete control and others assume continuous control. In this discourse we confine our attention to continuous control. Our main aims are to provide a timely review of the current state-of-the-art and to propose several important extensions to the existing theory. There is apparently no current review that attempts to collect together the most general form of the theory in a systematic way. For instance one must go to the closely-related theory for control of solar-powered racing cars in a paper by Howlett and Pudney (1998) to find the key ideas underpinning a fundamental existence result for optimal driving strategies.

The classic single train control problem is to minimise the energy required to drive a train from one station to the next within a given time. We assume continuous control. If no energy is recovered during braking then the optimal strategy is essentially a Maximum Power–HoldP–Coast–Maximum Brake strategy except that the singular HoldP (Hold using Power) phase with constant speed $v = V$ and positive control must be interrupted by phases of regular control to negotiate steep grades. One must insert phases of Maximum Power to traverse steep uphill sections and Coast with zero control to traverse steep downhill sections. Hence the optimal strategy becomes an optimal switching strategy. By reformulating the necessary conditions for optimal switching, Howlett et al. (2009) showed in a less general model that the optimal switching points for each steep section can be found by minimising an intrinsic local energy functional. If energy is recovered during braking then an optimal strategy may include additional singular HoldR (Hold using Regenerative braking) phases with the same constant speed $v = W > V$ and negative control on steep downhill sections. The analysis of optimal strategies involves examining the close relationship between the speed profile of the train and the profile of a modified adjoint variable—both expressed as functions of position. A representative optimal speed profile and corresponding modified adjoint profile are shown in Fig. 1.

The main contributions to calculation of optimal strategies for continuous control are due to Albrecht et al. (2013b), Howlett (2000), Howlett et al. (2009), Khmelnitsky (2000) and Liu and Golovitcher (2003). There are also important contributions to the fundamental theory. Howlett and Pudney (1998) used standard techniques of functional analysis to establish the existence of optimal driving strategies for solar-powered racing cars and have remarked elsewhere (Howlett, 2000) that a similar argument applies to trains subject to certain basic feasibility conditions. Albrecht et al. (2013b) and Khmelnitsky (2000) have used independent arguments to each propose conditions that ensure the optimal strategy is unique.

A key aim of this two-part paper is to show that all important results remain valid for a theoretically tractable model in which more general functional forms are allowed for both tractive and braking control and for resistance to motion. Despite a significant extension of the theory there are instances where the new results are more comprehensive. One such instance relates to the use

of regenerative braking. In Part 1 we pay particular attention to formulation of the model, analysis of the strategies of optimal type, location of optimal switching points and construction of phase plots (Vu, 2006) for evolutionary lines. In Part 2 (Albrecht et al., 2014) we prove existence of an optimal strategy, extend previous arguments to establish a local energy minimization principle for determination of optimal switching points (Albrecht et al., 2011; Howlett et al., 2009), and prove uniqueness of the optimal strategy with less restrictive assumptions (Albrecht et al., 2013b). To conclude we consider computation of optimal controls in two examples involving complex gradient profiles and discuss a typical speed profile for a complete optimal strategy.

1.1. Practical implementation of optimal driving strategies

The commercial *Energymiser*® Driver Advice System was invented by SCG researchers and developed in cooperation with Sydney-based company TTG Transportation Technology (TTG). The system minimises the mechanical energy required to drive the train and is used by railways around the world. A major part of the implementation is a precise route map of the track to ensure an accurate gradient profile. *Energymiser*® measures the location and speed of the train using GPS technology, calculates the optimal driving strategy to the next stop, and provides advice to the driver in real time. The in-cab display may simply be the current recommended control and subsequent control with an estimated time in seconds to the next control change, or it may be a more comprehensive display that also shows the corresponding speed profile. By repeatedly recalculating the optimal strategy, the system overcomes inaccuracies in the model parameters and state observations, and the effects of weather and train loading that cause the train to deviate from the predicted path. The advice is filtered to remove short phases that would result in the advice changing too frequently for a driver. The SCG and TTG are currently working with major train manufacturers to embed *Energymiser*® into automatic train operation systems which will enable the train to follow the ideal profile more closely. *Energymiser*® has improved time-keeping and has reduced fuel consumption by up to 20% in audited in-service operation.

1.2. A general model for optimal train control with position as the independent variable

Howlett and Pudney (1995) showed that motion of a train with distributed mass can be reduced to motion of a point-mass train. Thus we restrict our attention to point-mass trains. We understand that this leaves open the question of energy dissipation caused by in-train forces and any consequent impact on the effectiveness of particular driving strategies—but that is a separate research issue. See, for instance, Zhou et al. (2013) and Zhuan and Xia (2006). In general terms the problem is to drive a train from $x = 0$ to $x = X$ within some prescribed time T in such a way that energy consumption is minimised. It has been argued or assumed by all major contributors (Howlett, 2000; Khmelnitsky, 2000; Liu and Golovitcher, 2003) that in order to calculate the precise optimal strategy on non-level track it is convenient to formulate the model with position $x \in [0, X]$ as the independent variable and with time $t = t(x) \in [0, T]$ and speed $v = v(x) \in [0, \infty)$ as the dependent state variables.

The equations of motion are

$$t' = \frac{1}{v} \tag{1}$$

$$v' = \frac{u - r(v) + g(x)}{v} \tag{2}$$

where $(t, v) = (t(x), v(x))$ for $x \in [0, X]$ and where $u = u(x) \in (-\infty, \infty)$ is the known measurable control—the force per unit mass or acceleration. We have written $t' = dt/dx$ and $v' = dv/dx$. We assume $v(0) = v(X) = 0$, $v = v(x) > 0$ for all $x \in (0, X)$ and that $u = u(x)$ is bounded. In particular we assume there are two functions $U_-(v)$ and $U_+(v)$ with the following properties. We have $U_-[v(x)] \leq u(x) \leq U_+[v(x)]$ for each $x \in (0, X)$. The bounds $U_- = U_-(v) \in (-\infty, 0)$ and $U_+ = U_+(v) \in (0, \infty)$ for $v \in (0, \infty)$ are monotone functions with $U_-(v) \uparrow 0$ and $U_+(v) \downarrow 0$ as $v \uparrow \infty$. We suppose too, for each $\epsilon > 0$, there exists some constant $U'_\epsilon > 0$ such that $|U_-(v) - U_-(w)| \leq U'_\epsilon |v - w|$ and $|U_+(v) - U_+(w)| \leq U'_\epsilon |v - w|$ for all $v, w \geq \epsilon$. The functions U_- and U_+ define bounds for the maximum braking and driving forces per unit mass in a form that includes—as special cases—the specified bounds for a wide range of railway traction systems.

Remark 1. Some explanations are in order for the properties of the bounds. In Section 1.3 we use elementary physics to argue that if friction is neglected then $dv/dt = p/v$ where p is the tractive power per unit mass at the wheels. Therefore—in the absence of force limits at low speed—the maximum tractive acceleration is $U_+(v) = P/v$ which means that $U_+(v) \in (0, \infty)$ and $U_+(v) \downarrow 0$ as $v \uparrow \infty$. In practice such an idealised formula is likely to be only approximately true. By simply assuming that $U_+(v) \downarrow 0$ as $v \uparrow \infty$, we obtain a model that is more likely to remain valid in practice. If $U_+(v) = P/v$ then $U'_+(v) = -P/v^2$ and so $|U_+(v) - U_+(w)| \leq (P/\epsilon^2)|v - w|$ provided $v, w \geq \epsilon$. Thus our assumption of a Lipschitz property for $U_+(v)$ is simply a natural generalisation of an existing idealised property. Similar remarks apply to the maximum braking bound $U_-(v)$.

The function $r(v)$ is a general resistance per unit mass with no specific formula assumed. We define auxiliary functions $\varphi(v) = vr(v)$ and $\psi(v) = v^2 r'(v)$ and assume only that $\varphi(v)$ is strictly convex with $\varphi(v) \geq 0$ for $v \geq 0$ and $\varphi(v)/v \to \infty$ as $v \to \infty$. It follows that both $r(v)$ and $\psi(v)$ are non-negative and strictly increasing for $v \geq 0$. These properties capture the functional characteristics of the traditional quadratic formula—the so-called Davis formula (Davis et al., 1926)—that has been used in practice by the rail industry for many years.

Remark 2. The Davis formula for resistance takes the form $r(v) = r_0 + r_1 v + r_2 v^2$ where r_0, r_1 and r_2 are non-negative physical constants that may be determined approximately in practice (Vu, 2006) by measuring the difference between the applied tractive force and the net tractive force. The formula is based on elementary physical considerations but nevertheless may not be precisely realised in practice. If the Davis formula is used then $\varphi(v) = r_0 v + r_1 v^2 + r_2 v^3$ is a strictly convex function and $\psi(v) = v^2 r'(v) = r_1 v^2 + 2 r_2 v^3$ is a strictly increasing function. Hence our assumptions are a natural generalisation of an idealised physical property. The convexity of $\varphi(v)$ ensures that speedhold where possible is the most energy-efficient strategy. The monotonicity of $\psi(v)$ implies that the optimal driving speed is uniquely defined for each optimal journey. Thus the properties of $\varphi(v)$ and $\psi(v)$ capture the essential elements of the resistance function.

The function $g(x)$ is nominally the component of gravitational acceleration due to track gradient but in practice may also include additional position-dependent resistive forces due to track curvature. We will not consider details of such calculations. Suffice to say that resistance due to curvature is often modelled effectively by calculating an equivalent gradient acceleration. We describe the track as *steep uphill* at speed $v = V$ if $U_+(V) - r(V) + g(x) < 0$ so that the speed decreases even when the control is set to $u = U_+(v)$. The track is said to be *steep downhill* at speed $v = V$ if $-r(V) + g(x) > 0$ so that the speed increases even when the control is set to $u = 0$. See Howlett (2000), Howlett et al. (2009), Khmelnitsky (2000) and Liu and Golovitcher (2003) for further discussion. Unless otherwise stated we assume that $g(x)$ is smooth for $x \in [0, X]$. Let $G_- = \min_{x \in [0,X]} g(x)$ and $G_+ = \max_{x \in [0,X]} g(x)$. We will assume that for some $\epsilon > 0$ we have $U_+(v) - r(v) + G_- > 0$ for $0 \le v < \epsilon$ so that the train can maintain speed $v \ge \epsilon$ at any point and also—for theoretical convenience—that $U_-(v) - r(v) + G_+ < 0$ for $0 \le v \le V_{\max}$ so that the train can slow down at any point. The latter assumption may be violated in some practical situations but these are effectively handled by rail operators on an *ad hoc* basis using auxiliary braking systems and by imposing local operating rules. The cost of a control strategy is the net mechanical energy per unit mass required to move the train, given by

$$J = \int_0^X \left[\frac{(u + |u|)}{2} + \frac{\rho(u - |u|)}{2} \right] dx \tag{3}$$

where $\rho \in [0, 1)$ is the proportion of mechanical energy recovered during braking.

Remark 3. The optimisation minimises the mechanical work done by the traction system. This is not the same as minimising fuel usage or energy supply because efficiency varies with speed and tractive effort. Most traction systems have peak efficiency at high load during phases of maximum power. On some trains with multiple traction systems it is possible to run different traction motors at different levels. For example, tests on long-haul freight trains in Australia showed that it was more efficient during extended HoldP phases to turn off one of the three locomotives and run the other two at maximum power rather than running all three locomotives at two-thirds power. In practice minimising mechanical work done results in significant reductions in actual fuel or energy use.

Remark 4. Electric and diesel–electric trains can use the electric motors for traction or braking but energy generated by braking is not always recovered. In diesel–electric trains it is usually dissipated as heat in large resistor banks. In this case the regeneration efficiency is $\rho = 0$. With electric trains, the braking energy can be fed back into the track power supply if other loads can use or store the energy. Thus regeneration efficiency depends on the availability of other loads. The level of efficiency does not make a significant difference to the optimal strategy or to the mechanical energy required to drive the train and so the main benefit is the collective value of the recovered energy. Recent work on integrated scheduling shows that adjusting dwell-times on busy networks may enable significant energy savings (Gong et al., 2014; Li and Lo, 2014a, 2014b).

Speed limits may be imposed in the form $v \le v_m$ where $v_m = v_m(x) \in [0, V_{\max}]$ is a known bounded measurable function. The problem formulation with x as the independent variable has a particular theoretical advantage in that it partially separates the state variables and thereby enables us to solve (2) for $v = v(x)$ without knowledge of $t = t(x)$ which is subsequently determined directly from $v = v(x)$ using (1). The time constraint takes the form $t(X) \le T$. It is also possible to consider a seemingly more natural formulation of the problem with time as the independent variable. Although this formulation is equivalent to our formulation it is generally less convenient for mathematical analysis. We refer the reader to an earlier paper for further details (Howlett, 2000).

1.3. Previous models

The SCG have used both time (Howlett, 2000) and position (Albrecht et al., 2013b; Howlett, 1996; Howlett and Jiaxing, 1997; Howlett et al., 2009; Howlett and Leizarowitz, 2001; Howlett et al., 1994) as the independent variable. Although these models were developed specifically to describe the observed traction characteristics of a standard diesel–electric locomotive they are nevertheless based on fundamental physical principles and are generally applicable to all traction systems. The original idea was that a diesel–electric locomotive should deliver an approximately constant level of power per unit mass p_s that was directly proportional to the rate of fuel supply for each fixed throttle setting s. See (Howlett, 1996) for more details. In mathematical terms this meant that if friction and track gradient were neglected then there was a constant level p_s of tractive power per unit mass defined by the formula

$$\frac{d[v^2/2]}{dt} = p_s \iff \frac{dv}{dt} = \frac{p_s}{v}$$

for each value s of the throttle setting. In contrast, the braking mechanism was assumed to be a directly-applied negative acceleration. Hence the basic equation of motion became

$$\frac{dv}{dt} = \frac{p}{v} - q - r(v) + g(x) \iff \frac{dv}{dx} = \frac{p/v - q - r(v) + g(x)}{v}$$

where $p \in [0, P]$ was the power per unit mass under traction and $q \in [0, Q]$ was the braking force per unit mass.

The previous SCG models with $U_+(v) = P/v$ and $U_-(v) = -Q$ for $v > 0$ were special cases of the general model described in Section 1.2. In practice the tractive power delivered by electric and diesel–electric traction systems is limited at low speeds by the current supplied to the motor and by adhesion. Thus a more accurate model for maximum tractive acceleration would take the form $U_+(v) = P/\max\{v, v_0\}$ for some $v_0 > 0$. The formula for maximum braking acceleration is more complex but a similar expression in the form $U_-(v) = -Q/\max\{v, v_0\}$ could be used as a reasonable approximation. The precise formula for maximum tractive acceleration at very small speeds is not especially important to overall costs because the formula applies to only a tiny portion of the journey. For similar reasons the precise formula for maximum braking acceleration—whatever the speed—makes little difference to the overall strategy and to the net energy consumption.

The previous SCG train control models did not allow regenerative braking although both regenerative braking and the capture and storage of solar energy were included in an essentially similar model for optimal control of solar-powered racing cars. See Howlett and Pudney (1998). The early SCG models were primarily concerned with discrete control (Cheng et al., 1999; Howlett, 1996; 2000; Howlett and Leizarowitz, 2001; Howlett et al., 1994; Howlett and Pudney, 1995) but later models mainly use continuous control (Albrecht et al., 2013b; Howlett, 2000; Howlett et al., 2009; Howlett and Pudney, 1995; 1998). Collectively the SCG papers document a sound basis for modern train control theory. Numerical algorithms that calculate and continually update the optimal strategy during the course of a journey were first developed by Howlett and Pudney under an Australian Research Council (ARC) Linkage Grant during 1998–2000 for TTG Transportation Technology. These algorithms have since been further developed in the *Energymiser*® Driver Advice System which has been used extensively to reduce energy consumption on both freight and passenger trains in Australasia and Europe.

Two other frequently cited papers—Khmelnitsky (2000) and Liu and Golovitcher (2003)—have proposed models that closely resemble the models used in the SCG papers (Albrecht et al., 2013b; Howlett, 2000; Howlett et al., 2009; Howlett and Pudney, 1995; 1998) and the model considered here. Nevertheless there are many useful relationships that have not previously been explored that will be elaborated here and others that will be extended.

Khmelnitsky (2000) formulated the train control problem with kinetic energy and time as the dependent state variables and position as the independent variable. Although it is unusual to use kinetic energy rather than speed as the key state variable there are sound mathematical and engineering reasons to justify this. In particular the basic formula

$$\frac{dK}{dx} = \frac{d[v^2/2]}{dx} = \left(\frac{1}{v}\right)\frac{d[v^2/2]}{dt} = \frac{dv}{dt}$$

for the tractive force per unit mass on the train when friction and gradient are neglected has no singularity at $v = 0$. Thus the Khmelnitsky formulation avoids the mathematical difficulty of the (removable) singularity at $v = 0$ that arises from the formula $dv/dx = (1/v)dv/dt$ when speed v is used as the state variable. We note too that this formulation sits well with the overall objective to minimise energy consumption. Khmelnitsky assumes the more restrictive Davis formula

$$r(K) = r_0 + r_1\sqrt{2K} + 2r_2K \iff r(v) = r_0 + r_1v + r_2v^2$$

for the resistance function but the model is otherwise equivalent to the more general model described in this paper. He obtained a complete description of the optimal strategy and presented a numerical algorithm that can be used to find the associated speed profile. He also provided an indirect argument that the optimal strategy is uniquely defined. We will not review this argument but prefer to consider the direct argument for uniqueness proposed recently in Albrecht et al. (2013b).

The model used by Liu and Golovitcher (2003) is essentially the same as the more general model proposed in this paper except that they do not explicitly consider cost recovery through regenerative braking. Although the initial formulation is given in a time-dependent form it is immediately converted into a position-dependent problem. The mathematical optimisation is solved by minimising a Lagrangian functional

$$L = J + \kappa T = \int_0^X \left[u_+(x)U_+(v) + \frac{\kappa}{v}\right]dx$$

subject to the constraints $v(0) = v(X) = 0$, $v(x) > 0$ for $0 < x < X$ and

$$v' = \frac{u_+U_+(v) + u_-U_-(v) - r(v) + g(x)}{v}$$

where the power and brake controls are bounded by $0 \le u_+(x) \le 1$ and $0 \le u_-(x) \le 1$ for all $x \in [0, X]$ and where $\kappa > 0$ is a Lagrange multiplier for the overall journey time constraint

$$\int_0^X (1/v)dx \le T.$$

An earlier paper by Howlett and Leizarowitz (2001) used a similar formulation to find a general solution with chattering control when only a finite number of throttle settings is allowed. The results presented by Liu and Golovitcher (2003) describing

the optimal strategy essentially confirm the results obtained by Howlett (2000) and Khmelnitsky (2000) but with one notable additional contribution. In practice, track gradients—or equivalently track gradient accelerations—are tabulated by rail companies as piecewise-constant functions. For such gradients, on each segment of constant slope, Liu and Golovitcher obtained an algebraic expression for the key adjoint variable as a function of speed and thereby greatly simplified the process of calculating optimal switching points. Similar formulae for a modified adjoint variable were discovered independently by Howlett et al. and used in numerical algorithms for the *Energymiser*® system. This work was later published by Vu (2006) and Howlett et al. (2009). Liu and Golovitcher also developed a numerical algorithm for calculation of optimal strategies.

Remark 5. In this paper we consider optimal driving strategies for a single train with a given journey time and no interference from other trains. In practice it may be necessary to use suboptimal strategies for individual trains on a busy network in order to avoid disruptions to the schedules of other co-located trains and to maintain safe separation between trains travelling in the same direction on the same line. There is a substantial existing literature on optimal scheduling (Cordeau et al., 1998; Ghoseiria et al., 2004; Gong et al., 2014; Higgins et al., 1996; Kraay et al., 1991; Li and Lo, 2014a; 2014b; Zhou and Zhong, 2005) and on rudimentary methods to integrate scheduling and energy-efficient driving but there are still unsolved problems relating to optimal driving strategies for fleets of trains (Albrecht et al., 2013a).

2. Existence of an optimal strategy

It is important from a theoretical viewpoint to establish the existence of an optimal strategy. In other words it is necessary to show that the optimal control problem—to drive a train from one station to the next within a given time in such a way that energy consumption is minimised—is well-posed and has a solution. There are certain key points to consider. One must allow sufficient time for the journey to be completed. This leads to the concept of a minimum-time strategy. Intuitively one might reasonably expect this strategy to consist of only two phases—an initial phase of Maximum Power and a final phase of Maximum Brake. This is indeed the case.

In general—if the minimum-time journey is feasible—we may expect a large number of different feasible strategies. Ultimately it is necessary to show that we can always choose a strategy from the set of feasible strategies so that energy consumption is minimised. There are various technical problems that must be addressed before this can be done. Firstly we must construct a set $\mathcal{F}$ of feasible control strategies—a set where each nominated control $u = u(x) \in \mathcal{F}$ generates uniquely-defined solutions $v = v(x)$ to (2) and $t = t(x)$ to (1) that satisfy all relevant state constraints. Secondly we must find a special sequence $\{S_n\} \in \mathcal{F}$ of feasible control strategies for which the energy consumption progressively decreases and in the limit approaches the least possible value. For any particular non-optimal strategy $S_n \in \mathcal{F}$ with $J(S_n) = \inf_{S \in \mathcal{F}} J(S) + \epsilon_n$ where $\epsilon_n > 0$ it follows from the definition of an infimum that there exists an improved strategy $S_{n+1} \in \mathcal{F}$ such that $J(S_{n+1}) < \inf_{S \in \mathcal{F}} J(S) + \epsilon_n/2$ and so the required special sequence exists by induction. Thirdly—and most importantly—we must show that the feasible set is sequentially compact. See Yosida (1978, The Eberlein-Šmulian theorem, pp. 141–145) for more information. That is, we must be able to extract a subsequence of control strategies from our special sequence that converges to a well-defined optimal strategy in the set $\mathcal{F}$. It is convenient to introduce these ideas here but postpone our discussion of the underlying mathematics. Details of the full mathematical argument will be presented in Part 2 of this paper (Albrecht et al., 2014). A similar proof that an optimal strategy exists for a solar-powered racing car can be found in Howlett and Pudney (1998).

2.1. The minimum-time strategy

We define two important speed profiles. The speed profile $v = v_P(x)$ for $x \geq 0$ is for the Maximum Power phase that starts at $(x, v) = (0, 0)$. The speed profile $v = v_B(x)$ for $x \leq X$ is for the Maximum Brake phase that finishes at $(x, v) = (X, 0)$. In mathematical parlance(Birkhoff and Rota (1989)), we define $v = v_P(x)$ as the unique solution for $x \geq 0$ to the equation

$$v' = \frac{U_+(v) - r(v) + g(x)}{v}$$

with $v_P(0) = 0$ and define $v = v_B(x)$ as the unique solution for $x \leq X$ to the equation

$$v' = \frac{[U_-(v) - r(v) + g(x)]}{v}$$

with $v_B(X) = 0$. The minimum-time speed profile $\bar{v} = \bar{v}(x) = \min[v_P(x), v_B(x)]$ for each $x \in [0, X]$ with $T_{\min} = \int_0^X dx/\bar{v}(x)$ is generated by a Maximum Power–Maximum Brake strategy with control $\bar{u}(x) = \bar{u}(x, \bar{v}(x))$ where

$$\bar{u}(x, v) = \begin{cases} U_+(v) & \text{for } x \in (0, c) \\ U_-(v) & \text{for } x \in (c, X) \end{cases}$$

and $c \in (0, X)$ is the unique point defined by $v_P(c) = v_B(c)$. A typical speed profile for the minimum-time strategy is depicted as the upper profile in Fig. 2. The minimum-time strategy is also a maximum-speed strategy. For every strategy with $v(0) = v(X) = 0$ we must have $v(x) \leq \bar{v}(x)$ for all $x \in [0, X]$.
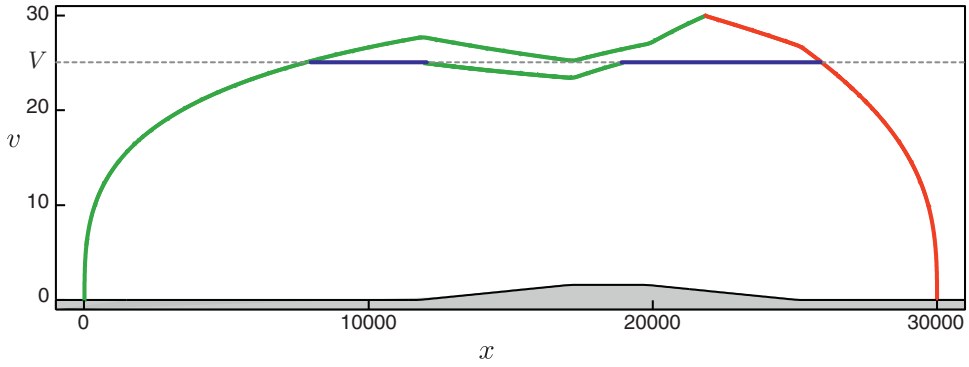
Fig. 2. Speed profiles $v = \bar{v}(x)$ for a typical minimum-time strategy (upper profile) and $v = \bar{v}_V(x)$ with $V = 25$ for a corresponding feasible approximate-speedhold strategy (lower profile). The position x is measured in metres (m) and the speed v is measured in metres per second (m s^{-1}).

2.2. Feasible strategies

A feasible control strategy is one in which the journey is completed within the prescribed time T. To pose a meaningful optimal control problem we need to know that an optimal strategy exists. A fundamental requirement is that the feasible control set be non-empty. Thus it is important to establish the existence of at least one feasible strategy. If $T_{\min} \leq T$ then the minimum-time strategy described in the previous subsection is a feasible strategy. If $T_{\min} < T$ then we can reduce energy consumption by reducing speed and increasing the journey time. A natural way to do this is to follow a constant maximum speed wherever possible—the first intuitive step towards constructing an optimal strategy. We proceed as follows.

Choose V with $0 < V < V_{\max} = \max_{x \in [0,X]} \bar{v}(x)$ where $v = \bar{v}(x)$ is the speed profile for the minimum-time strategy. Define a control function

$$u_V(x, v) = \begin{cases} U_+(v) & \text{if } v < V \\ r(V) - g(x) & \text{if } v = V \end{cases}$$

and a corresponding speed profile $v = v_V(x)$ for $x \in [0, X]$ satisfying the differential equation

$$vv' = u_V(x, v) - r(v) + g(x)$$

with $v_V(0) = V$. This is an approximate-speedhold strategy on $[0, X]$ starting at speed V. The idea is that if the speed falls below $v = V$ then we apply Maximum Power. However, because $v_V(0) = V \neq 0$, this strategy is not feasible. Provided the nominal hold speed V is sufficiently high we can construct a corresponding feasible approximate-speedhold strategy with speed profile

$$\bar{v}_V(x) = \min[\bar{v}(x), v_V(x)]$$

and control function

$$\bar{u}_V(x, v) = \begin{cases} \bar{u}(x) & \text{if } \bar{v}_V(x) = \bar{v}(x) \\ u_V(x, v) & \text{if } \bar{v}_V(x) = v_V(x) \end{cases}$$

where $v = \bar{v}(x)$ is the speed profile for the minimum-time strategy. The time taken for this strategy will be $T_V = \int_0^X dx/\bar{v}_V(x)$. The strategy will be feasible if $T_{\min} < T_V \leq T$. In general, if $T_{\min} < T$, there will be an infinite collection of feasible strategies. A typical profile for an approximate-speedhold strategy is shown as the lower profile in Fig. 2.

2.3. Speedhold is the most cost-efficient form of control

The approximate-speedhold strategy proposed in the previous subsection is not optimal but it is nevertheless true that speedhold is an energy-efficient driving mode. Suppose we consider a speedhold phase with driving control $u(x, V) = r(V) - g(x) > 0$ at constant speed $v(x) = V$ for all $x \in [p, q]$. The time taken is given by

$$\Delta T_V = \int_p^q \frac{dx}{V} = \frac{q - p}{V}$$

and the cost is

$$\Delta J_V = \int_p^q [r(V) - g(x)]dx = r(V)(q - p) - G(q) + G(p).$$

For any alternative strategy with control $u(x) > 0$ and $v(p) = v(q) = V$ we can integrate the identity $v(x)v'(x) = u(x) - r[v(x)] + g(x)$ to give

$$0 = \int_p^q u(x)dx - \int_p^q r[v(x)]dx + G(q) - G(p) \iff \Delta J_v = \int_p^q r[v(x)]dx - G(q) + G(p)$$

where ΔJ_v is the cost of the alternative strategy. Since $\varphi(v) = vr(v)$ is strictly convex we have $\varphi(v) > L_V(v) = \varphi(V) + \varphi'(V)(v - V)$ for all $v \neq V$ and hence $r(v) - r(V) > Vr'(V)(1 - V/v)$ for $v \neq V$. It follows that

$$\Delta J_v - \Delta J_V = \int_p^q \{r[v(x)] - r(V)\}dx > Vr'(V)\int_p^q [1 - V/v(x)]dx \tag{4}$$

provided $v(x) \neq V$ for some $x \in (p, q)$. If the alternative strategy with $v = v(x)$ takes the same time to traverse $[p, q]$ as the speedhold strategy then $\int_p^q dx/v(x) = (q - p)/V$ and the inequality (4) becomes $\Delta J_v - \Delta J_V > 0$. Thus a speedhold phase is more cost efficient than any alternative strategy of positive control.

3. The Hamiltonian analysis

This is the main section of the paper and it is appropriate to summarise the content and provide a rationale for the way in which the material is organised. We begin by introducing the Hamiltonian function and the associated adjoint variables in Section 3.1. The Pontryagin principle is applied in Section 3.2 to show that only certain special control modes can be used in an optimal strategy and that the choice of control is determined by the values of the speed and the corresponding values of the adjoint variables. In order to understand how an optimal strategy evolves it is necessary to follow the joint evolution of these variables. The general principles for this evolution are discussed in Section 3.3. The evolution is a complex process and it is convenient to begin by considering the special case of a track with constant gradient. We also look at several different methods of analysis.

In Sections 3.4 and 3.5 we consider evolution of the state and adjoint variables on track with constant gradient. In this case the adjoint variables can be expressed as algebraic functions of the speed. The relevant formulae are derived in Section 3.5 and are then used in Section 3.6 to find a general form for the strategy of optimal type on level track. These formulae also allow us to construct phase diagrams for the speed and the adjoint variables that show a complete range of possible evolutionary lines for a strategy of optimal type. The phase diagrams are described in Section 3.6 for level track and in Section 3.7 for steep uphill and steep downhill track. In practice rail operators tabulate track gradients as piecewise-constant functions of position. Since the values of the speed and the corresponding values of the adjoint variables determine the level of control it is important for effective on-board calculations that the algebraic formulae from Section 3.5 are extended to track with piecewise-constant gradient. This is done in Section 3.8.

In Sections 3.9, 3.10 and 3.11 we consider the full range of possible control transitions at the end of each regular phase of optimal control. In Section 3.12 we introduce the problem of finding optimal switching points when a singular control phase is interrupted by a regular control phase to negotiate a steep section of track or alternatively when a change in gradient means it is necessary to change the level of singular control. In Section 3.13 we show that active speed limits are essentially an imposed form of singular control.

It is timely to reiterate that the main purpose of this two-part paper is to show that many results that were previously known to apply in special circumstances remain true with a more general model. In particular our analysis in this section allows us to identify the critical factors that support many of these familiar results. For instance the magnitude bounds on the tractive and braking forces must decrease as the speed increases in order to preserve the characteristic properties of the evolutionary lines in the phase plane; the auxiliary function $\varphi(v)$ must be strictly convex to ensure that a constant-speed strategy is the most energy-efficient; and the auxiliary function $\psi(v)$ must be strictly increasing to establish the existence of a unique optimal driving speed for each optimal journey. Thus we show that the optimal driving speed is the same for every HoldP phase and that the optimal regenerative braking speed is the same for every HoldR phase in an optimal strategy. In the course of exploring these key ideas we elaborate various alternative techniques that can be used to describe, determine and analyse an optimal strategy. We discuss evolution of optimal strategies and pay particular attention to control transitions. This section also paves the way for a more general discussion in Part 2 of the paper (Albrecht et al., 2014) about the local energy functional introduced in Howlett et al. (2009) and the perturbation analysis proposed in Albrecht et al. (2013b).

3.1. The Hamiltonian function

To find necessary conditions on an optimal strategy we use a Hamiltonian analysis and apply the Pontryagin principle (Girsanov et al., 1972; Hartl et al., 1995). We use the model defined by (1), (2) and (3) with position as the independent variable. An equivalent analysis is possible with time as the independent variable but the details are different (Howlett, 2000). In the first instance we assume there are no imposed speed limits. Define the Hamiltonian function

$$\mathcal{H} = (-1)\left[\frac{(u + |u|)}{2} + \frac{\rho(u - |u|)}{2}\right] + \frac{\mu_1}{v} + \frac{\mu_2[u - r(v) + g(x)]}{v} \tag{5}$$

and the associated Lagrangian function

$$\mathcal{L} = \mathcal{H} + \pi_1(U_+(v) - u) + \pi_2(u - U_-(v)) \tag{6}$$

where $\pi_1 \geq 0$ and $\pi_2 \geq 0$ are Lagrange multipliers for the control constraints. The adjoint variables $(\mu_1, \mu_2) = (\mu_1(x), \mu_2(x))$ are absolutely continuous and satisfy the differential equations

$$\mu_1' = (-1)\frac{\partial \mathcal{L}}{\partial t} = 0 \tag{7}$$

$$\mu_2' = (-1)\frac{\partial \mathcal{L}}{\partial v} = \frac{\mu_1}{v^2} + \frac{\mu_2[u - r(v) + g(x)]}{v^2} + \frac{\mu_2 r'(v)}{v} - \pi_1 U_+'(v) + \pi_2 U_-'(v). \tag{8}$$

3.2. Necessary conditions on an optimal strategy

For an optimal strategy the Pontryagin principle shows it is necessary to find a control that maximises the Hamiltonian at each point subject to the given constraints. Thus we set

$$\frac{\partial \mathcal{L}}{\partial u} = (-1)\left[\frac{(1 + \text{sgn}(u))}{2} + \frac{\rho(1 - \text{sgn}(u))}{2}\right] + \frac{\mu_2}{v} - \pi_1 + \pi_2 = 0 \tag{9}$$

which gives $-1 + \mu_2/v - \pi_1 + \pi_2 = 0$ when $u > 0$ and $-\rho + \mu_2/v - \pi_1 + \pi_2 = 0$ when $u < 0$. When $u = 0$ the partial derivative in (9) is undefined. It is also necessary to satisfy the Karush–Kuhn–Tucker conditions $\pi_1(U_+(v) - u) = 0$ and $\pi_2(u - U_-(v)) = 0$. A strategy that maximises the Hamiltonian at each point will be called a *strategy of optimal type*. Before beginning a detailed analysis we note from (7) that $\mu_1 = b$ for some constant b. There are five possible modes.

Mode 1: $\mu_2 > v$. We must have $u = U_+(v) > 0$, $\pi_1 = \mu_2/v - 1 > 0$ and $\pi_2 = 0$. This is a *regular* control phase of Maximum Power.

Mode 2: $\mu_2 = v$. We must have $u > 0$, $\pi_1 = 0$ and $\pi_2 = 0$. If the condition is maintained over a nontrivial interval $x \in (p, q)$ then $\mu_2' = v' \Rightarrow \psi(v) + b = 0$. Since $\psi(v)$ is strictly increasing in v this equation has only one solution $v = V = V_b > 0$ where $b = -\psi(V) < 0$. We call V the *optimal driving speed*. This is a *singular* control phase denoted by HoldP with hold speed $v(x) = V$ and positive power $0 < u(x) = r(V) - g(x) < U_+(V)$ for $x \in (p, q)$.

Mode 3: $v > \mu_2 > \rho v$. We must have $u = 0$, $\pi_1 = 0$ and $\pi_2 = 0$. This is a *regular* control phase of Coast.

Mode 4: $\mu_2 = \rho v$. We must have $u < 0$, $\pi_1 = 0$ and $\pi_2 = 0$. If the condition is maintained over a nontrivial interval $x \in (p, q)$ then $\mu_2' = \rho v' \Rightarrow \rho \psi(v) + b = 0$. Since $\psi(v)$ is strictly increasing in v this equation has only one solution $v = W = W_{\rho, b} > 0$ where $b = -\rho\psi(W) < 0$. We call W the *optimal regenerative braking speed*. This is a *singular* control phase denoted by HoldR with hold speed $v(x) = W$ and negative power $U_-(W) < u(x) = r(W) - g(x) < 0$ for $x \in (p, q)$. Note that $0 \leq \rho < 1$ and so $\rho\psi(W) = \psi(V) \Leftrightarrow W = \psi^{-1}[\psi(v)/\rho]$ means that $0 < V < W$. If we allow $\rho = 1$ then we obtain $W = V$.

Mode 5: $\rho v > \mu_2$. We must have $u = U_-(v) < 0$, $\pi_1 = 0$ and $\pi_2 = \rho - \mu_2/v > 0$. This is a *regular* control phase of Maximum Brake.

For each mode of optimal control the Lagrange multipliers are given by

$$\pi_1 = \frac{|\mu_2/v - 1| + (\mu_2/v - 1)}{2} = \frac{|\eta| + \eta}{2} = \begin{cases} \eta & \text{if } \eta > 0 \\ 0 & \text{if } \eta \leq 0 \end{cases} \tag{10}$$

where we have defined a modified adjoint variable $\eta = \mu_2/v - 1$ and

$$\pi_2 = \frac{|\mu_2/v - \rho| - (\mu_2/v - \rho)}{2} = \frac{|\zeta| - \zeta}{2} = \begin{cases} 0 & \text{if } \zeta \geq 0 \\ -\zeta & \text{if } \zeta < 0 \end{cases} \tag{11}$$

where we have defined an alternative modified adjoint variable $\zeta = \mu_2/v - \rho = \eta + 1 - \rho$.

3.3. Evolution of a strategy of optimal type

For each strategy of optimal type $\mu_1(x) = \mu_1 = -\psi(V) = -\rho\psi(W)$ is a constant that defines a unique optimal driving speed $v = V$ with $u(x) = r(V) - g(x) > 0$ and a unique optimal regenerative braking speed $v = W = \psi^{-1}[\psi(V)/\rho] > V$ with $u(x) = r(W) - g(x) < 0$. The optimal driving speed $v = V$ can only be maintained on non-steep track where $g(x) \leq r(V) \leq U_+(V) + g(x)$. The optimal regenerative braking speed $v = W$ can only be maintained on steep downhill track where $U_-(W) + g(x) \leq r(W) \leq g(x)$.

Evolution of the singular control phases HoldP and HoldR is straightforward. In each case both the state variable $v = v(x)$ and adjoint variable $\mu_2 = \mu_2(x)$ are constant. Hence it is possible to exit a singular control phase in a strategy of optimal type at any point and still remain optimal. However we shall see later that there is only one exit point that will find a feasible strategy. The collective length of all singular phases in a strategy of optimal type is determined by the overall distance constraints.

Evolution of the regular control phases in a strategy of optimal type is more complex and is determined in the following way. Let $x = p$ be a point on a strategy of optimal type. Select a permissible control $u = U_+(v)$ (if $v(p) < \mu_2(p)$), $u = 0$ (if $0 < \mu_2(p) < v(p)$) or $u = U_-(v)$ (if $\mu_2(p) < 0$) and solve (2) to find $v = v(x)$ on some proposed interval $[p, q]$ subject to a given initial speed $v = v(p)$. Once $v = v(x)$ is known solve (8) on $[p, q]$ to find $\mu_2 = \mu_2(x)$ subject to the given initial value $\mu_2 = \mu_2(p)$. If the values of $(v(x), \mu_2(x))$ on (p, q) are consistent with optimality for the selected control $u = u(v)$ then the combined profile $(x, v, \mu_2) = (x, v(x), \mu_2(x))$ for $x \in (p, q)$ is an allowable phase in a strategy of optimal type. The decision to exit a regular control phase is completely determined by the evolution of $(v(x), \mu_2(x))$.

In general (2) must be solved by numerical integration for each phase of regular optimal control. Rather than solving (8) for $\mu_2 = \mu_2(x)$ to study possible control transitions it is convenient to use one or other of the modified adjoint variables, $\eta = \mu_2/v - 1$ or $\zeta = \mu_2/v - \rho = \eta + 1 - \rho$, introduced earlier in this section. We may use (2), (8), (10) and (11) to show that

for each strategy of optimal type, the modified adjoint variables satisfy linear differential equations written in standard form[1] as

$$\eta' - \frac{\psi(v) - v^2 u'(v)}{v^3}\eta = \frac{\psi(v) - \psi(V)}{v^3} \tag{12}$$

for $\eta = \eta(x)$ in the region $\mu_2 > \rho v \Leftrightarrow \eta > \rho - 1 \Leftrightarrow \zeta > 0$ with $u(v) = U_+(v)$ for $\eta > 0$ and $u(v) = 0$ for $\rho - 1 < \eta < 0$ and

$$\zeta' - \frac{\psi(v) - v^2 u'(v)}{v^3}\zeta = \rho\frac{\psi(v) - \psi(W)}{v^3} \tag{13}$$

for $\zeta = \zeta(x)$ in the region $\mu_2 < v \Leftrightarrow \zeta < 1 - \rho \Leftrightarrow \eta < 0$ with $u(v) = 0$ for $0 < \zeta < 1 - \rho$ and $u(v) = U_-(v)$ for $\zeta < 0$ where, in (12) and (13), $v = v(x)$ is the known solution to (2) with the appropriate level of control. Note that both (12) and (13) are valid in the region $\rho - 1 < \eta < 0 \Leftrightarrow 0 < \zeta < 1 - \rho$ and in this region it is simply a matter of preference as to which equation and which variable is used. If we wish to find an analytic expression for the solution to (12) on an interval $[p, q]$ then we can define an integrating factor

$$I_{p,u}(x) = \exp\left[-\int_p^x \frac{\psi(v) - v^2 u'(v)}{v^3}d\xi\right] \tag{14}$$

for all $x \in [p, q]$ where the known speed profile has been written as $v = v(\xi)$ for convenience in the integrand. If we multiply both sides of (12) by the integrating factor and integrate over $[p, x] \subset [p, q]$ we obtain

$$I_{p,u}(x)\eta(x) - \eta(p) = \int_p^x \frac{\psi(v) - \psi(V)}{v^3}I_{p,u}(\xi)d\xi \tag{15}$$

for all $x \in [p, q]$ provided either $\eta(x) > 0$ with $u(v) = U_+(v)$ for all $x \in (p, q)$ or $\rho - 1 < \eta(x) < 0$ with $u(v) = 0$ for all $x \in (p, q)$. A similar argument using (13) shows that

$$I_{p,u}(x)\zeta(x) - \zeta(p) = \rho\int_p^x \frac{\psi(v) - \psi(W)}{v^3}I_{p,u}(\xi)d\xi \tag{16}$$

for all $x \in [p, q]$ provided either $0 < \zeta(x) < 1 - \rho$ with $u(v) = 0$ for all $x \in (p, q)$ or $\zeta(x) < 0$ with $u(v) = U_-(v)$ for all $x \in (p, q)$. Once again, in both (15) and (16), we have written $v = v(\xi)$ for the known speed profile.

We can find linear differential equations[2] for $\eta = \eta(x)$ in the region $\eta < \rho - 1$ and for $\zeta = \zeta(x)$ in the region $\zeta > 1 - \rho$ but in each case the form of the equation is less convenient. Thus it is simpler to use the relationship $\eta(x) = \zeta(x) + \rho - 1$ to determine $\eta(x)$ when $\eta(x) < \rho - 1 \Leftrightarrow \zeta(x) < 0$ and to use the relationship $\zeta(x) = \eta(x) + 1 - \rho$ to find $\zeta(x)$ when $\zeta(x) > 1 - \rho \Leftrightarrow \eta(x) > 0$.

The broader problem is to construct a strategy of optimal type consisting entirely of permissible optimal control phases in such a way that the combined profiles $(x, v, \eta) = (x, v(x), \eta(x))$ or $(x, v, \zeta) = (x, v(x), \zeta(x))$ are continuous at all linking points. In general this is a complicated process but we will begin by considering an important special case where the gradient acceleration is constant on the selected track segment and evolution of both the state and modified adjoint variables is more easily described.

3.4. Evolution of the state variables on track with constant gradient

Suppose a segment of track (p, q) has constant gradient acceleration $g(x) = \gamma$ for all $x \in (p, q)$. We will show that for each permissible phase of regular control in a strategy of optimal type there is an analytic solution to the equations of motion.

1. **Maximum Power**—with $\eta > 0$ and control $u(v) = U_+(v)$. Since $U_+(v) \downarrow 0$ is strictly decreasing and $r(v)\uparrow\infty$ is strictly increasing as $v\uparrow\infty$ there is a unique limiting speed $V_+ = V_{+,\gamma}$ under Maximum Power defined by solving the equation $U_+(v) - r(v) + \gamma = 0$. We may separate the variables and integrate (2) to obtain the analytic solution. For $v(p) < V_+$ we have

$$x(v) - p = \int_{v(p)}^v \frac{w\,dw}{U_+(w) - r(w) + \gamma}, \quad t(v) - t(p) = \int_{v(p)}^v \frac{dw}{U_+(w) - r(w) + \gamma} \tag{17}$$

for all $v \in (v(p), V_+)$ with $x, t\uparrow\infty$ as $v \uparrow V_+$. For $v(p) > V_+$ we have

$$x(v) - p = \int_v^{v(p)} \frac{-w\,dw}{U_+(w) - r(w) + \gamma}, \quad t(v) - t(p) = \int_v^{v(p)} \frac{-dw}{U_+(w) - r(w) + \gamma} \tag{18}$$

for all $v \in (V_+, v(p))$ with $x, t\uparrow\infty$ as $v \downarrow V_+$. In each case we have a strictly monotonic solution $x = f_+(v)$ for position as a function of speed with a well-defined inverse function $v = f_+^{-1}(x)$. For $v(p) < V_+$ we have $v \uparrow V_+$ as $x\uparrow\infty$ and for $v(p) > V_+$ we have $v \downarrow V_+$ as $x\uparrow\infty$. Thus, if the phase is maintained indefinitely, the train approaches a limiting Maximum Power speed $v = V_+$ for the given gradient.

---

[1] We use the expression $\psi(v)/v^3$ in (12) and thereafter because it turns out to be more convenient than the seemingly simpler $r'(v)/v$. In particular the fact that $\psi(v)$ is strictly increasing in v means the ubiquitous terms $[\psi(v) - \psi(V)]/v^3$ and $[\psi(v) - \psi(W)]/v^3$ are respectively negative when $v < V$ or $v < W$ and positive when $v > V$ or $v > W$.

[2] For $\eta < \rho - 1$ we have $\eta' - \{[\psi(v) - v^2 U'_-(v)]/v^3\}\eta = [\psi(v) - \psi(V)]/v^3 - (1 - \rho)U'_-(v)/v$ and for $\zeta > 1 - \rho$ we have $\zeta' - \{[\psi(v) - v^2 U'_+(v)]/v^3\}\zeta = \rho[\psi(v) - \psi(W)]/v^3 + (1 - \rho)U'_+(v)/v$.

2. **Coast**—with $\rho - 1 < \eta < 0 \Leftrightarrow 0 < \zeta < 1 - \rho$ and control $u(v) = 0$. Suppose that $\gamma > r(0) \Leftrightarrow -r(0) + \gamma > 0$. Since $r(v) \uparrow \infty$ is strictly increasing as $v \uparrow \infty$ there exists a unique limiting speed $V_0 = V_{0, \gamma}$ under Coast defined by solving the equation $-r(v) + \gamma = 0$. We may separate the variables and integrate (2) to obtain the analytic solution. For $v(p) < V_0$ we have

$$x(v) - p = \int_{v(p)}^{v} \frac{wdw}{-r(w) + \gamma}, \quad t(v) - t(p) = \int_{v(p)}^{v} \frac{dw}{-r(w) + \gamma} \tag{19}$$

for all $v \in (v(p), V_0)$ with $x, t \uparrow \infty$ as $v \uparrow V_0$. For $v(p) > V_0$ we have

$$x(v) - p = \int_{v}^{v(p)} \frac{-wdw}{-r(w) + \gamma}, \quad t(v) - t(p) = \int_{v}^{v(p)} \frac{-dw}{-r(w) + \gamma} \tag{20}$$

for all $v \in (V_0, v(p))$ with $x, t \uparrow \infty$ as $v \downarrow V_0$. In each case we have a strictly monotonic solution $x = f_0(v)$ for position as a function of speed with a well-defined inverse function $v = f_0^{-1}(x)$. For $v(p) < V_0$ we have $v \uparrow V_0$ as $x \uparrow \infty$ and for $v(p) > V_0$ we have $v \downarrow V_0$ as $x \uparrow \infty$. Thus, if the phase is maintained indefinitely, the speed approaches the limiting Coast speed $v = V_0$ for the given gradient. If $\gamma < r(0)$ then for all $v(p) > 0$ we have

$$x(v) - p = \int_{v}^{v(p)} \frac{-wdw}{-r(w) + \gamma} < \int_{0}^{v(p)} \frac{-wdw}{-r(w) + \gamma} < \infty \tag{21}$$

and

$$t(v) - t(p) = \int_{v}^{v(p)} \frac{-dw}{-r(w) + \gamma} < \int_{0}^{v(p)} \frac{-dw}{-r(w) + \gamma} < \infty \tag{22}$$

for all $v \in (0, v(p))$. Hence we have a monotonic decreasing solution $x = f_0(v)$ for position as a function of speed with a well-defined inverse function $v = f_0^{-1}(x)$ but in this case there is some finite point $q > p$ with $v(q) = 0$. In this case Coast will stop the train after a finite time.

3. **Maximum Brake**—with $\zeta < 0$ and control $u(v) = U_-(v)$. We assume that $\gamma < -U_-(v) + r(v) \Leftrightarrow U_-(v) - r(v) + \gamma < 0$ for all $v > 0$. We may separate the variables and integrate (2) to obtain the analytic solution. For each $v(p) > 0$ we have

$$x(v) - p = \int_{v}^{v(p)} \frac{-wdw}{U_-(w) - r(w) + \gamma} < \int_{0}^{v(p)} \frac{-wdw}{U_-(w) - r(w) + \gamma} < \infty \tag{23}$$

and

$$t(v) - t(p) = \int_{v}^{v(p)} \frac{-dw}{U_-(w) - r(w) + \gamma} < \int_{0}^{v(p)} \frac{-dw}{U_-(w) - r(w) + \gamma} < \infty \tag{24}$$

for all $v \in (0, v(p))$. Hence we have a monotonic decreasing solution $x = f_-(v)$ for position as a function of speed with a well-defined inverse function $v = f_-^{-1}(x)$ and some finite point $q > p$ with $v(q) = 0$. Thus Maximum Brake will stop the train after a finite time.

3.5. Evolution of the adjoint variables on track with constant gradient

An important practical observation first published by Liu and Golovitcher (2003) and later developed by Vu (2006) and Howlett et al. (2009) is that on track with constant gradient, during phases of regular control in a strategy of optimal type, the adjoint variables can be determined as algebraic functions of speed alone. These formulae are particularly important as they can be extended by continuity to provide algebraic formulae for the adjoint variables on track with piecewise-constant gradient. Track gradients, or equivalently track gradient accelerations, are recorded by railways as piecewise-constant functions. Hence the algebraic formulae can be—and have been—used effectively for fast on-board calculations to determine switching points in strategies of optimal type. We will derive analogous expressions for our more general model. Before we begin it is convenient to define a key function $E_Z : (0, \infty) \to [\varphi'(Z), \infty)$ by the formula

$$E_Z(v) = \frac{\psi(Z)}{v} + r(v). \tag{25}$$

Note that

$$E_Z'(v) = -\frac{\psi(Z)}{v^2} + r'(v) = \frac{\psi(v) - \psi(Z)}{v^2}$$

which is negative when $v < Z$ and positive when $v > Z$. Hence $E_Z : (0, \infty) \to [\varphi'(Z), \infty)$ is strictly quasiconvex[3] and has a unique minimum turning point at $v = Z$ with $E_Z(Z) = \varphi'(Z)$. A typical graph $y = E_Z(v)$ with $Z = 25$ and $r(v) = 1.0 \times 10^{-2} + 1.5 \times 10^{-5}v^2$ is shown in Fig. 3. The speed v is measured in metres per second (m s^{-1}) and the resistance $r(v)$ and key function $E_Z(v)$ are measured in metres per second squared (m s^{-2}).

---

[3] The function $f : (0, \infty) \to \mathbb{R}$ is said to be *strictly quasiconvex* if $f(\alpha x + (1 - \alpha)y) < \max\{f(x), f(y)\}$ for all $x, y \in (0, \infty)$ and all $\alpha \in (0, 1)$.
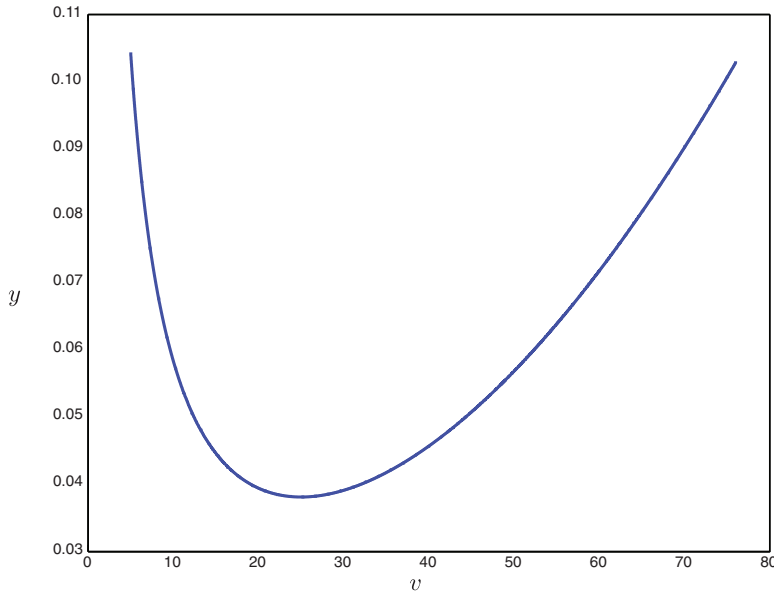
Fig. 3. Typical graph $y = E_Z(v)$ for $Z = 25$. The speed v is measured in metres per second (m s^{-1}) and the function $y = E_Z(v)$ is measured in metres per second squared (m s^{-2}).

Let V and W denote the respective optimal driving and optimal regenerative braking speeds. Once again we consider a permissible phase of regular control in a strategy of optimal type on an interval $[p, q]$ where the gradient acceleration is constant. That is, $g(x) = \gamma$ for all $x \in (p, q)$. We will show that in each case there are algebraic formulae for the modified adjoint variables as functions of speed only.

1. **Maximum Power**—with $\eta > 0$ and control $u(v) = U_+(v)$. From (2) we have

$$\frac{d}{dx}[vv'] = [U'_+(v) - r'(v)]v' = (-1)\frac{\psi(v) - v^2 U'_+(v)}{v^3}[vv'].$$ (26)

It follows from (12) and (26) that

$$\frac{d}{dx}\{\eta[vv']\} = \left\{\eta' - \frac{\psi(v) - v^2 U'_+(v)}{v^3}\eta\right\}vv' = \left\{\frac{\psi(v) - \psi(V)}{v^2}\right\}v' = E'_V(v)\frac{dv}{dx}$$

and hence $\eta[vv'] = E_V(v) + E$ where E is a constant of integration. For $v \neq V_{+,\gamma}$ we deduce that

$$\eta(v) = \frac{E_V(v) + E}{U_+(v) - r(v) + \gamma}.$$ (27)

2. **Coast**—with $\rho - 1 < \eta < 0 \Leftrightarrow 0 < \zeta < 1 - \rho$ and control $u(v) = 0$. For phases starting or finishing at $\eta = 0 \Leftrightarrow \zeta = 1 - \rho$ it is convenient to find a formula for $\eta = \eta(v)$. From (2) we have

$$\frac{d}{dx}[vv'] = -r'(v)v' = (-1)\frac{\psi(v)}{v^3}[vv'].$$ (28)
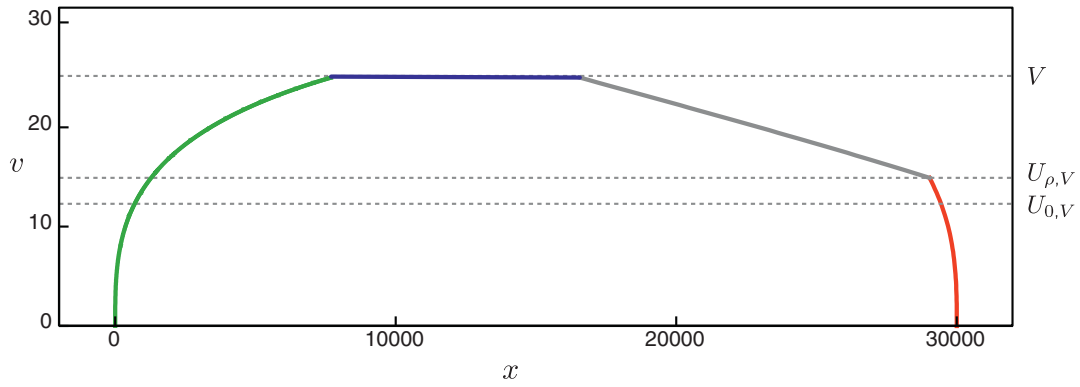
It follows from (12) and (28) that

$$\frac{d}{dx}\{\eta[vv']\} = \left\{\eta' - \frac{\psi(v)}{v^3}\eta\right\}vv' = \left\{\frac{\psi(v) - \psi(V)}{v^2}\right\}v' = E'_V(v)\frac{dv}{dx}$$

and hence $\eta[vv'] = E_V(v) + E$ where E is a constant of integration. For $v \neq V_{0,\gamma}$ we have

$$\eta(v) = \frac{E_V(v) + E}{-r(v) + \gamma}.$$ (29)

For phases starting or finishing at $\zeta = 0 \Leftrightarrow \eta = \rho - 1$ it is convenient to find a formula for $\zeta = \zeta(v)$. It follows from (13) and (28) that

$$\frac{d}{dx}\{\zeta[vv']\} = \left\{\zeta' - \frac{\psi(v)}{v^3}\zeta\right\}vv' = \rho\left\{\frac{\psi(v) - \psi(W)}{v^2}\right\}v' = \rho E'_W(v)\frac{dv}{dx}$$

Fig. 4. Optimal speed profile on level track for $\rho > 0$ with braking speed $U_{\rho, V} \in (U_{0, V}, V)$. The position x is measured in metres (m) and the speed v is measured in metres per second (m s^{-1}).

and hence $\zeta[vv'] = \rho E_W(v) + F$ where F is a constant of integration. For $v \neq V_{0, \gamma}$ it follows that

$$\zeta(v) = \frac{\rho E_W(v) + F}{-r(v) + \gamma}. \tag{30}$$

3. **Maximum Brake** with $\zeta < 0$ and control $u(v) = U_-(v)$. A similar argument shows there is some constant F with

$$\zeta(v) = \frac{\rho E_W(v) + F}{U_-(v) - r(v) + \gamma}. \tag{31}$$

For an optimal phase on a segment (p, q) with constant gradient acceleration $g(x) = \gamma$ it is important to reiterate that although the speed $v = v(x)$ depends on x, the adjoint variable $\eta = \eta(v) = \eta_\gamma(v)$ depends only on v. This means that for each optimal driving speed and each constant gradient acceleration we can construct universal phase plots $(v, \eta(v))$ for the optimal trajectories.[4] We will refer to individual trajectories as *evolutionary lines*. The phase plots are independent of x but we can calculate the distance travelled and time taken between any two points (v_1, η_1) and (v_2, η_2) on the same evolutionary line using the formulae in Section 3.4.

We will use phase plots of the evolutionary lines to explain various feasible control transitions. In practice—where rail operators tabulate track gradient accelerations on (x_p, x_q) as piecewise-constant functions—an optimal phase plot will consist of a finite sequence of linked evolutionary line segments $A_{j-1}A_j$ from (v_{j-1}, η_{j-1}) to (v_j, η_j) with constant gradient accelerations $g(x) = \gamma_j$ on each segment (x_{j-1}, x_j) for $j = p+1, \ldots, q$. In general, with non-constant gradient, the evolutionary lines for $(x, v(x), \eta(x))$ also depend on the position x but at any particular point $x = p$ we can study local behaviour by looking at evolutionary lines for the constant gradient acceleration $\gamma = g(p)$. It may also be convenient, on some occasions, to simply plot the projected evolutionary lines $(v(x), \eta(x))$ in the phase plane.

### 3.6. The strategy of optimal type on level track

We have argued that every strategy of optimal type is determined by the evolution of the state and adjoint variables along an optimal evolutionary line $(x, v(x), \eta(x))$. On track with piecewise-constant gradient we showed that local evolution is independent of position. That is, for constant gradient $g(x) = \gamma$, we have $\eta = \eta(v) = \eta_\gamma(v)$ and the evolutionary lines take the local form $(v, \eta(v))$. We will use this information to show that on level track the optimal strategy is a Maximum Power–HoldP–Coast–Maximum Brake strategy. A typical optimal speed profile for level track is shown in Fig. 4.

The optimal strategy on level track consists of a linked sequence of evolutionary lines on the phase plot for level track. A typical phase plot is shown in Fig. 5. To construct this plot we used $U_+(v) = P/\max\{v_0, v\}$, $U_-(v) = Q/\max\{v_0, v\}$ with $P = 1$, $Q = -1$ and $v_0 = 5$, $r(v) = r_0 + r_2 v^2$ with $r_0 = 1.0 \times 10^{-2}$ and $r_2 = 1.5 \times 10^{-5}$ and $\rho = 0.5$. The nominated optimal driving speed is $V = 25$ and the corresponding speed at which braking begins is $U_{\rho, V} \approx 14.90$. The limiting speed under Maximum Power is obtained by solving the equation $P/\max\{v_0, v\} - r(v) + \gamma = 0$ with $\gamma = 0$ to give $v = V_{+, \gamma} = V_{+, 0} \approx 35.11$.

To construct an optimal strategy on level track consider the possible evolutionary lines for (v, η). Since $v = 0$ at $x = 0$ the initial phase must be a phase of Maximum Power with $\eta > 0$ and control $u = U_+(v)$. Hence the initial phase on the evolutionary line is defined by the formula

$$\eta(v) = \frac{E_V(v) + E}{U_+(v) - r(v)}$$

---

[4] Since $\zeta = \eta + 1 - \rho$ it follows that equivalent phase plots for $(v, \zeta(v))$ are simply a translation of the phase plots for $(v, \eta(v))$. In general we will restrict our attention to plots for $(v, \eta(v))$.
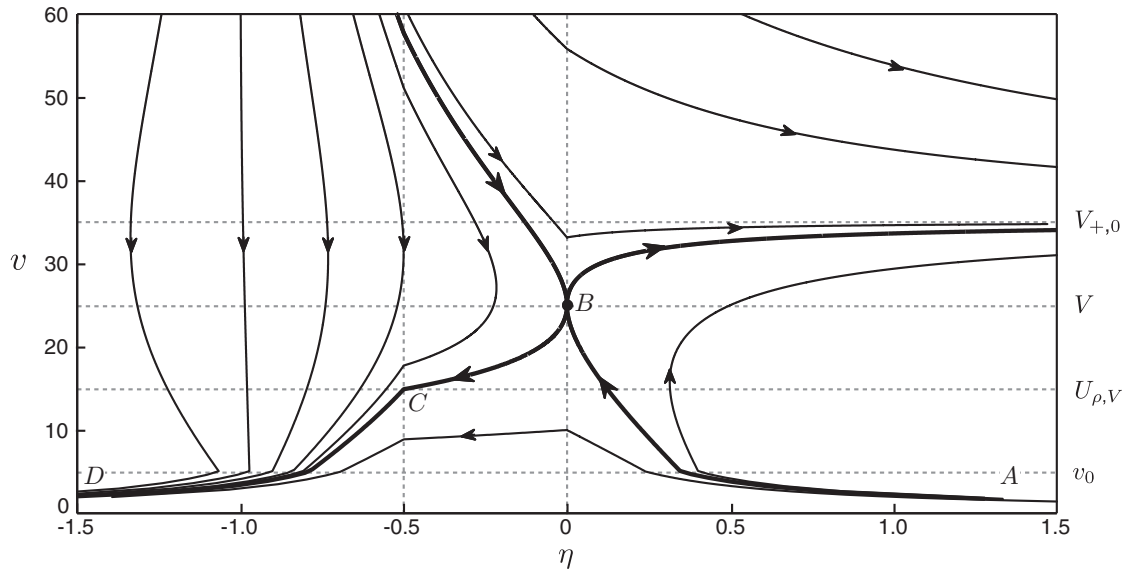
Fig. 5. Phase plots for evolutionary lines on level track. The lines in the region $\eta \geq 0$ are Maximum Power curves, the lines in the region $-0.5 \leq \eta \leq 0$ are Coast curves and the lines in the region $\eta < -0.5$ are Maximum Brake curves. The modified adjoint variable η is dimensionless and the speed v is measured in metres per second (m s^{-1}).

where $V > 0$ is the nominated optimal driving speed and E is an unknown constant. Since $U_+(v) - r(v) > 0$ when v is small it follows from (2) that $v = v(x)$ increases as x increases. If Maximum Power is maintained *ad infinitum* then $v(x) \uparrow V_{+,0}$ as $x \uparrow \infty$. However, in a strategy of optimal type this is not feasible and so this segment of the evolutionary line must terminate at some point B with $x = b < X$. The necessary conditions for optimality show that $(v(b), \eta(b)) = (V, 0)$ at the termination point B and so $E = -E_V(V) = -\varphi'(V)$. Therefore

$$\eta(v) = \frac{E_V(v) - E_V(V)}{U_+(v) - r(v)}$$

which is zero at $v = V$ but is otherwise positive for $0 < v < V_{+,0}$. Thus $\eta = \eta(v)$ has a minimum turning point with $(v, \eta) = (V, 0)$ at B. The Maximum Power phase is represented by the evolutionary line AB in Fig. 5 but we note that the initial point A with coordinates $(v, \eta) = (0, \eta(0))$ on the phase plot may be the point $(0, \infty)$. The distance travelled along AB is given by

$$\Delta x_{AB} = \int_0^V \frac{v dv}{U_+(v) - r(v)} \iff b = \int_0^V \frac{v dv}{U_+(v) - r(v)}.$$

We may now switch to HoldP with control $u = r(V)$. If so then the entire singular HoldP phase is represented by the point B in Fig. 5. The phase may be maintained as long as we please[5] but at some stage we must switch to Coast with $\rho - 1 < \eta < 0 \Leftrightarrow 0 < \zeta < 1 - \rho$ and control $u(v) = 0$. During Coast the speed decreases and the modified adjoint variable $\eta = \eta(v)$ is given by the formula

$$\eta(v) = \frac{E_V(v) - E_V(V)}{-r(v)}$$

which is zero if $v = V$ and is otherwise negative. Hence $\eta = \eta(v)$ has a maximum turning point $(v, \eta) = (V, 0)$ at B. The Coast phase is represented by the evolutionary line BC in Fig. 5 and will terminate when $\eta = \rho - 1$. Therefore the speed at this point is given by solving

$$\rho - 1 = \frac{E_V(v) - E_V(V)}{-r(v)} \iff \rho \varphi(v) = L_V(v) \tag{32}$$

where $L_V(v) = \varphi'(V)(v - V) + \varphi(V)$. The graph $y = L_V(v)$ is the straight line tangent to the strictly convex curve $y = \varphi(v)$ at the point $v = V$. The tangent line intersects the v-axis at

$$U_{0,V} = V - \frac{\varphi(V)}{\varphi'(V)} = \frac{\psi(V)}{\varphi'(V)} > 0. \tag{33}$$

---

[5] It is possible to have a HoldP phase of zero length which essentially means a direct switch from Maximum Power to Coast. We shall see shortly that the length of the HoldP phase must be adjusted so that the total distance travelled is correct. The length of the regular phases cannot be adjusted.
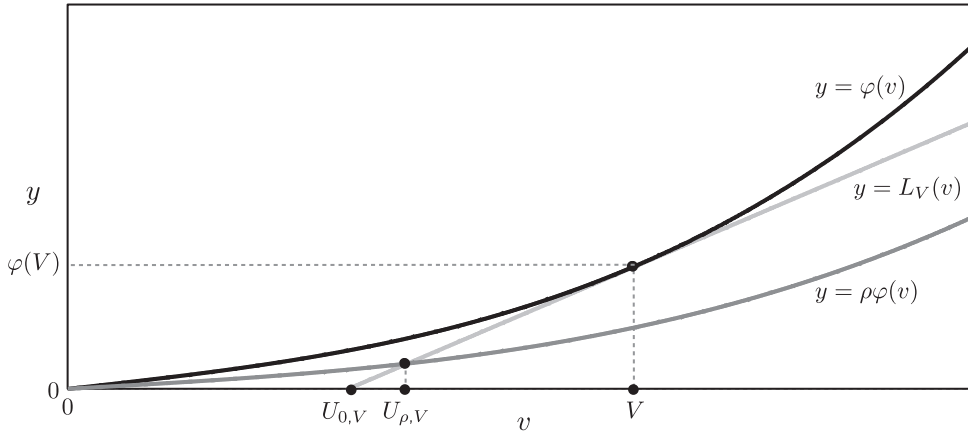
Fig. 6. The curves $y = \varphi(v)$, $y = L_V(v)$ and $y = \rho\varphi(v)$ showing the solution to the brake speed equation on level track with $0 < \rho < 1$ and $U_{0,V} < U_{\rho,V} < V$. The speed v is measured in metres per second (m s^{-1}) and the functions $y = \varphi(v)$, $y = L_V(v)$ and $y = \rho\varphi(v)$ are measured in metres squared per second cubed (m^2 s^{-3}).

The graph $y = \rho\,\varphi(v) > 0$ is strictly convex and lies below the graph $y = \varphi(v)$. Hence there is a unique solution $v = U_{\rho,V}$ with $U_{0,V} \leq U_{\rho,V} < V$ to (32). The speed $U = U_{\rho,V}$ is the speed at which braking begins[6] and it is uniquely defined by the optimal driving speed V and the efficiency ρ. If $\rho = 0$ then $U = U_{0,V}$. If $\rho = 1$ then $U = V$. A geometric interpretation of the solution to (32) is shown in Fig. 6.

The distance travelled during the Coast phase is given by

$$\Delta x_{BC} = \int_U^V \frac{v\,dv}{r(v)} \iff c - b = \int_U^V \frac{v\,dv}{r(v)}.$$

Although the algebra is somewhat tedious we can show that for the Coast phase we have

$$\eta(v) = \frac{E_V(v) - E_V(V)}{-r(v)} \iff \zeta(v) = \rho \frac{E_W(v) - E_W(U_{\rho,V})}{-r(v)}.$$

In purely algebraic terms the transition from Coast to Maximum Brake is managed more easily using $\zeta(v)$ rather than $\eta(v)$ because the switch occurs at $\eta = \rho - 1 \Leftrightarrow \zeta = 0$. In any event we switch to a Maximum Brake phase with $u = U_-(v)$ at point C in Fig. 5 where $v = U_{\rho,V}$ and $\eta = \rho - 1 \Leftrightarrow \zeta = 0$. The speed now decreases to $v = 0$ and the journey is complete. This final phase is represented by the evolutionary line CD in Fig. 5. During the Maximum Brake phase we have $\eta(v) = \zeta(v) + \rho - 1$ where

$$\zeta(v) = \rho \frac{E_W(v) - E_W(U_{\rho,V})}{U_-(v) - r(v)}$$

in the region $0 < v < U_{\rho,V}$. The distance travelled during the Maximum Brake phase is given by

$$\Delta x_{CD} = \int_0^{U_{\rho,V}} \frac{-v\,dv}{U_-(v) - r(v)} \iff d - c = \int_0^{U_{\rho,V}} \frac{-v\,dv}{U_-(v) - r(v)}.$$

The total distance travelled is X and so the distance travelled during the HoldP phase must be $\Delta x_B = X - (\Delta x_{AB} + \Delta x_{BC} + \Delta x_{CD})$. The time taken for the journey is given by

$$t(X) = \int_0^V \frac{dv}{U_+(v) - r(v)} + \frac{\Delta x_B}{V} + \int_{U_{\rho,V}}^V \frac{-dv}{-r(v)} + \int_0^{U_{\rho,V}} \frac{-dv}{U_-(v) - r(v)}.$$

If the time taken is too long—the time constraint $t(X) \leq T$ is not satisfied—we must increase the nominal value of V. If the time taken is too short, we can decrease the cost by decreasing the nominal value of V and increasing the time taken. The correct optimal journey can be found by iterating on the choice of V. See Cheng and Howlett (1992) and Howlett and Cheng (1993) for more details.

If $X < \Delta x_{AB} + \Delta x_{BC} + \Delta x_{CD}$ then we choose an evolutionary line underneath the curve $ABCD$ which does not reach the desired optimal driving speed and switches directly from Maximum Power to Coast at some nominal switching speed $V' < V$. Calculation of the distance travelled and time taken is similar but the HoldP phase is omitted. For the nominated value of V an iteration on V' will allow us to find the correct distance. We can then calculate the time taken. If this is not correct we select another value of V and repeat the procedure. Since the speed at which braking begins $U_{\rho,V}$ depends on V it follows that the distance and time

---

[6] It is not possible to switch to HoldR at $v = U$ because $U < V < W$.
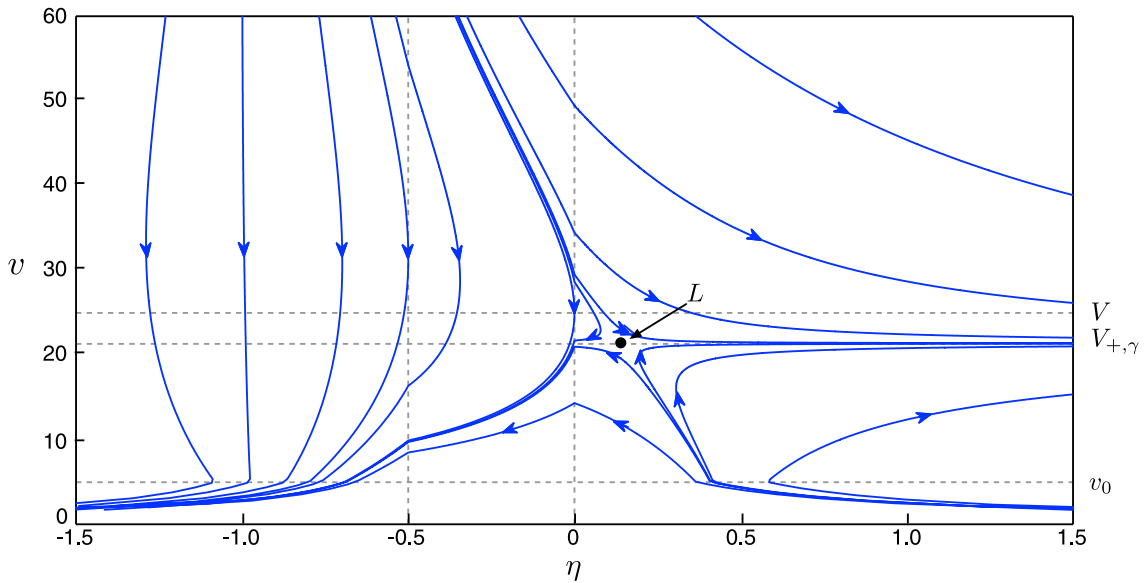
Fig. 7. Phase plots for evolutionary lines on track with constant gradient acceleration $\gamma = -0.03$ that is steep uphill at speed $V > V_{+,\gamma}$. The lines in the region $\eta \geq 0$ are Maximum Power curves, the lines in the region $-0.5 \leq \eta \leq 0$ are Coast curves and the lines in the region $\eta < -0.5$ are Maximum Brake curves. The modified adjoint variable η is dimensionless and the speed v is measured in metres per second (m s^{-1}).

calculations both depend on V and V'. During the Coast phase we have $v \leq V'$ and $\eta(v) = (-1)[E_V(v) - E_V(V')]/r(v)$. Therefore the speed $v = U < V'$ at which braking begins is obtained by solving

$$\eta(U) = \rho - 1 \iff U = V'\left[1 - \frac{\varphi(V') - \rho\,\varphi(U)}{\psi(V) + \varphi(V')}\right].$$

Since $U < V'$ it follows that $\rho\varphi(U) < \varphi(V')$ and hence that $U \uparrow V'$ as $V \uparrow \infty$. Thus the minimum-time journey is obtained by taking the limit as $V \uparrow \infty$. When $T = T_{\min}$ the feasible set contains only one strategy and the Pontryagin principle does not apply directly. This allows the adjoint variable $\eta(v)$ to have a jump discontinuity at $v = V'$.

3.7. Phase plots for strategies of optimal type on steep track with constant gradient

For a track with constant gradient acceleration $g(x) = \gamma$ and a given optimal driving speed we can use the algebraic formulae for the modified adjoint variables $\eta = \eta(v)$ to draw a phase plot in the (v, η) plane that shows all possible evolutionary lines for regular control phases in a strategy of optimal type. Figs. 7 and 8 respectively show typical phase plots for evolutionary lines on track with $\gamma = -0.03$ that is steep uphill at speed $V < V_{+,\gamma}$ and on track with $\gamma = 0.03$ that is steep downhill at speed $V > V_{0,\gamma}$. We have used the same train model and the same designated optimal driving speed $V = 25$ for all three phase plots of the evolutionary lines in Figs. 5, 7 and 8.

For the phase plot depicted in Fig. 7 we solve the equation $P/\max\{v_0, v\} - r(v) + \gamma = 0$ with $\gamma = -0.03$ to find the limiting speed $v = V_{+,\gamma} = V_{+,-0.03} \approx 21.35$ under Maximum Power. Since $V_{+,\gamma} < V$ we know that this track is steep uphill at speed $V = 25$ and indeed that any speed $v \geq V_{+,\gamma}$ cannot be maintained under Maximum Power. Fig. 7 shows phase curves under Maximum Power in the form

$$\eta(v) = \frac{E_V(v) + E}{P/\max\{v, v_0\} - r(v) + \gamma}$$

where $E \neq -E_V(V_{+,\gamma})$ with various initial points $(v(0), \eta(0))$ in the region $\eta \geq 0$. For $E > -E_V(V_{+,\gamma})$ the curves all have limit point $(V_{+,\gamma}, \infty) \approx (21.35, \infty)$ which cannot be reached in finite time. The point L in Fig. 7 with coordinates given by $v = V_{+,\gamma}$ and $\eta = E'_V(V_{+,\gamma})/[U'_+(V_{+,\gamma}) - r'(V_{+,\gamma})]$ is a limiting point that lies on the evolutionary line with $\eta(v) = [E_V(v) - E_V(V_{+,\gamma})]/[U_+(v) - r(v) + \gamma]$. This point also cannot be reached in finite time. For $E < -E_V(V_{+,\gamma})$ the curves reach the line $\eta = 0$ with $v < V_{+,\gamma}$ in finite time and transform into Coast curves. Fig. 7 also shows Coast curves in the region $-0.5 = \rho - 1 \leq \eta \leq 0$ in the form

$$\eta(v) = \frac{E_V(v) + E}{-r(v) + \gamma}$$

and Maximum Brake curves in the region $\eta \leq \rho - 1 = -0.5$. Note that if $E > -E_V(V)$ and $v(0) > V$ then the Coast curves reach the line $\eta = 0$ with $v > V$ and transform into Power curves.

For the phase plot depicted in Fig. 8 there are two important limiting speeds. We solve the equation $P/\max\{v_0, v\} - r(v) + \gamma = 0$ with $\gamma = 0.03$ to find the limiting speed $v = V_{+,\gamma} = V_{+,0.03} \approx 51.31$ under Maximum Power and solve the equation $-r(v) + \gamma =$
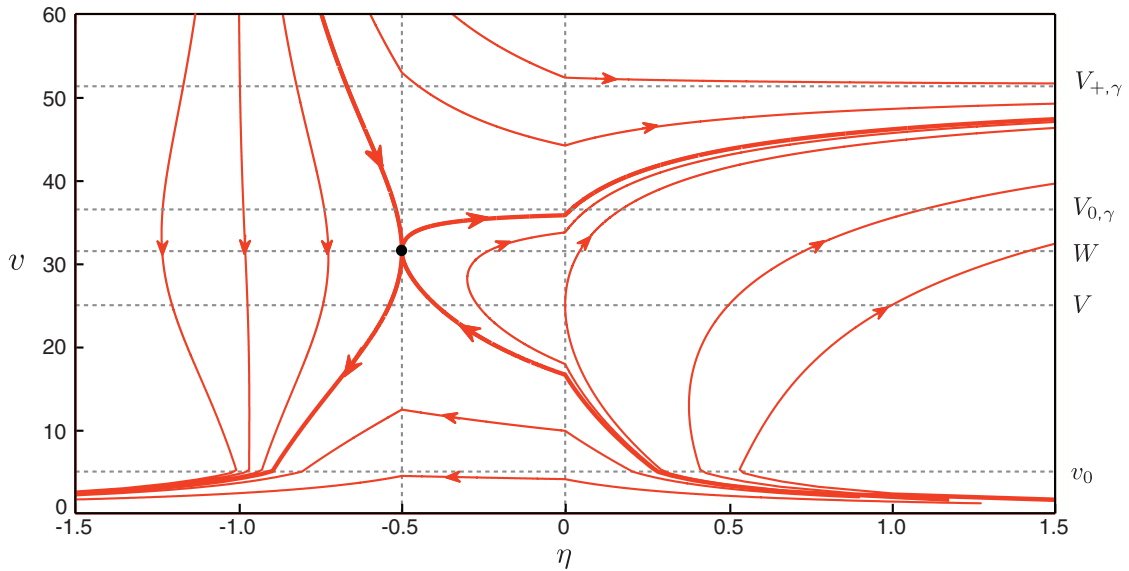
Fig. 8. Phase plots for evolutionary lines on track with constant gradient acceleration $\gamma = 0.03$ that is steep downhill at speed $V < V_{0,\gamma}$. The lines in the region $\eta \geq 0$ are Maximum Power curves, the lines in the region $-0.5 \leq \eta \leq 0$ are Coast curves and the lines in the region $\eta < -0.5$ are Maximum Brake curves. The modified adjoint variable η is dimensionless and the speed v is measured in metres per second (m s^{-1}).

0 with $\gamma = 0.03$ to find the limiting speed $v = V_{0,\gamma} = V_{0,0.03} \approx 36.51$ during Coast. Since $V < V_{0,\gamma}$ the track is steep downhill at speed V and the speed will increase during Coast if $v < V_{0,\gamma}$. Fig. 8 shows Power curves with various initial points $(v(0), \eta(0))$ in the region $\eta \geq 0$. If $v(0) = 0$ and $E \geq -E_V(V)$ then the evolutionary lines remain as Power curves forever. If $v(0) = 0$ and $-E_V(V) > E \geq -\rho E_W(W) + \gamma(\rho - 1) = -0.5E_W(W) - 0.5\gamma$ then the evolutionary lines reach the line $\eta = 0$ with $v < V$ and transform into Coast curves in the region $-0.5 < \eta < 0$ before returning to the line $\eta = 0$ with $V < v < V_{0,\gamma}$ and transforming back into Maximum Power curves. Fig. 8 also shows Maximum Brake curves in the region $\eta < -0.5$.

There are many things we can learn about optimal driving strategies from the phase plots of the evolutionary lines in Figs. 5, 7 and 8 but there are also complications. Firstly there is a different phase plot for every different optimal driving speed and every different gradient—although it must be said that the plots change continuously with continuous changes in the optimal driving speed and gradient. Secondly it is true that for most tracks, the gradient continually changes as we move along the track. On track with piecewise-constant gradient we must move the point $(v, \eta) = (v(x), \eta(x))$ for $x \in (x_{j-1}, x_j)$ along a segment of the evolutionary line using a plot for the current gradient $g(x) = \gamma_j$ until we reach the point $x = x_j$ where the gradient changes. At this stage it is necessary to move the point $(v, \eta) = (v(x), \eta(x))$ for $x \in (x_j, x_{j+1})$ along a new segment of the evolutionary line using a plot for the new gradient $g(x) = \gamma_{j+1}$.

3.8. Evolution of the adjoint variables on track with piecewise-constant gradient

Now suppose $x_p < x_{p+1} < \cdots < x_q$ with $g(x) = \gamma_j$ if $x \in (x_{j-1}, x_j)$ for each $j = p+1, \ldots, q$. We say that the gradient is piecewise-constant[7] on (x_p, x_q). For a regular control phase with either $\eta(v) > 0$ and $u(v) = U_+(v)$ or with $\rho - 1 < \eta(v) < 0$ and $u(v) = 0$ we have

$$\eta(v) = \frac{E_V(v) + E_j}{u(v) - r(v) + \gamma_j} \tag{34}$$

where $v = v(x)$ for each $x \in (x_{j-1}, x_j)$. If we write $v_j = v(x_j)$, $\eta_j = \eta(v_j)$ then we can use the continuity of $\eta = \eta(v)$ at $x = x_j$ to deduce from the left-hand limit that $[u(v_j) - r(v_j) + \gamma_j]\eta_j = E_V(v_j) + E_j$ and from the right-hand limit that $[u(v_j) - r(v_j) + \gamma_{j+1}]\eta_j = E_V(v_j) + E_{j+1}$. Thus it follows that

$$E_{j+1} - E_j = (\gamma_{j+1} - \gamma_j)\eta_j$$

for each $j = p+1, \ldots, q-1$. If $v_p = v_q = V$ and $\eta(V) = 0$ then $E_p = E_q = -E_V(V)$. For a regular control phase on (x_p, x_q) with either $0 < \zeta(v) < 1 - \rho$ and $u(v) = 0$ or with $\zeta(v) < 0$ and $u(v) = U_-(v)$ we have

$$\zeta(v) = \frac{\rho E_W(v) + F_j}{u(v) - r(v) + \gamma_j} \tag{35}$$

---

[7] Although it could be argued that in practice track gradients must be continuous it is nevertheless true that rail track data is tabulated in this form.

where $v = v(x)$ for each $x \in (x_{j-1}, x_j)$. If we write $\zeta = \zeta(v_j)$ and use the continuity of $\zeta = \zeta(v)$ at $x = x_j$ then a similar argument shows that

$$F_{j+1} - F_j = (\gamma_{j+1} - \gamma_j)\zeta_j$$

for each $j = p + 1, \ldots, q - 1$. If $v_p = v_q = W$ and $\zeta(W) = 0$ then $F_p = F_q = -\rho E_W(W)$.

3.9. Ending an optimal phase of Maximum Power

Let V be the optimal driving speed. For each optimal phase of Maximum Power we have $\eta > 0$ and $u(v) = U_+(v)$. Both $v = v(x)$ and $\eta = \eta(x)$ are everywhere continuous. If a phase of Maximum Power terminates at $x = q$ then $\eta(q) = 0$. We make the following specific observations.

1. Since $u'(v) = U'_+(v) < 0$ and $\eta > 0$ for $x < q$ it follows from (12) that $\eta' > \psi(v)\eta/v^3 > 0$ when $v = v(x) > V$. Hence Maximum Power can never terminate at a point $x = q$ with $v(q) > V$.
2. Maximum Power can terminate at $x = q$ with $\eta(q) = 0$ and $v(q) = V$ by changing to HoldP at the optimal driving speed $v = V$. In this case (12) shows that $\eta'(q) = 0$ at the termination point.
3. Maximum Power can terminate at $x = q$ with $\eta(q) = 0$ and $v(q) \leq V$ by changing to a Coast phase with $\eta(x) < 0$ and $v(x) < V$ when $x > q$. If $v(q) < V$ at the transition point with $\eta(q) = 0$ then (12) shows that $\eta'(q) = [\psi(v) - \psi(V)]/v^3 < 0$. Since $v = v(x)$ is continuous we cannot change to HoldP with $v(x) = V$ for $x > q$ if $v(q) < V$. Hence we must change to Coast.

When $u(v) = U_+(v)$ we write $I_{p,u}(x) = I_{p,+}(x)$ for the integrating factor defined in (14). With these substitutions (15) gives an analytic formula for $\eta = \eta(x)$ during a phase of Maximum Power. Although it is convenient to formulate the solution as if we were solving (12) in the forward direction it is better to compute a numerical solution in the backward direction. To explain this observation we note that the strict convexity of $\varphi(v)$ implies $\varphi'(v) > \varphi(v)/v = r(v)$ is increasing and so $r'(v) > 0$. Note also that $u'(v) = U'_+(v) < 0$. Therefore the exponent in (14) is negative and $I_{p,+}(x)$ becomes very small when $x \gg p$. Thus $\eta(x)$ is not accurately defined in numerical terms by (15) when $x \gg p$.

3.10. Ending an optimal phase of Coast

Let V and W denote the optimal driving and optimal regenerative braking speeds respectively. For each optimal phase of Coast we have $\rho - 1 < \eta < 0 \Leftrightarrow 0 < \zeta < 1 - \rho$ and $u(v) = 0$. If the phase terminates at $x = q$ then either $\eta(q) = 0 \Leftrightarrow \zeta(q) = 1 - \rho$ or else $\eta(q) = \rho - 1 \Leftrightarrow \zeta(q) = 0$. We make the following specific observations about ending Coast with $\eta = 0$.

1. Since $u'(v) = 0$ and $\rho - 1 < \eta < 0$ for $x < q$ it follows from (12) that $\eta' < \psi(v)\eta/v^3 < 0$ for $v < V$. Hence η decreases and so Coast cannot terminate at $x = q$ with $\eta(q) = 0$ when $v(q) < V$.
2. If $v > V$ when $x < q$ then Coast can terminate at $x = q$ with $\eta(q) = 0$ and $v(q) = V$ by changing to HoldP at the optimal driving speed $v = V$ for $x > q$. In this case (12) shows that $\eta'(q) = 0$ at the termination point.
3. Coast can terminate at $x = q$ with $\eta(q) = 0$ and $v(q) \geq V$ by changing to a phase of Maximum Power with $\eta > 0$ and $v > V$ when $x > q$. If $v(q) > V$ at the termination point with $\eta(q) = 0$ then (12) shows that $\eta'(q) = [\psi(v) - \psi(V)]/v^3 > 0$. Since $v = v(x)$ is continuous we cannot change to HoldP with $v = V$ for $x > q$ if $v(q) > V$. Hence we must change to Maximum Power.

We make the following specific observations about ending Coast with $\zeta = 0$.

1. Since $u'(v) = 0$ and $0 < \zeta < 1 - \rho$ for $x < q$ we deduce from (13) that $\zeta' > \psi(v)\zeta/v^3 > 0$ when $v > W$ and so ζ increases. Hence this phase cannot terminate at $x = q$ with $\zeta(q) = 0$ when $v(q) > W$.
2. The phase can terminate at $x = q$ with $\zeta(q) = 0$ and $v(q) = W$ by changing to HoldR at the optimal regenerative braking speed $v = W$ for $x > q$. In this case (13) shows that $\zeta'(q) = 0$ at the termination point.
3. Coast can terminate at $x = q$ with $\zeta(q) = 0$ and $v(q) \leq W$ by changing to Maximum Brake. If $v(q) < W$ then (13) shows that $\zeta'(q) < 0$ at the termination point. Since $v = v(x)$ is continuous we cannot change to HoldR with $v(x) = W$ for $x > q$ if $v(q) < W$. Hence we must change to Maximum Brake.

When $u(v) = 0$ we write $I_{p,u}(x) = I_{p,0}(x)$ for the integrating factor defined in (14). With these substitutions (15) and (16) give analytic formulae for $\eta = \eta(x)$ and $\zeta = \zeta(x)$ respectively during a phase of Coast. The exponent in (14) is negative and so, once again, it is better to compute numerical solutions in the backward direction.

3.11. Ending an optimal phase of Maximum Brake

Let W be the optimal regenerative braking speed. For each optimal phase of Maximum Brake we have $\zeta < 0$ and $u(v) = U_-(v)$. We assume that the speed always decreases when $u(v) = U_-(v)$. We make the following observations.

1. Suppose there is some point $x = p$ where $v(p) = W - \epsilon$ for some $\epsilon > 0$ and $\zeta(p) < 0$. As long as Maximum Brake continues we must have $v < W - \epsilon$. Let us choose δ so that $0 < \delta < |\zeta(p)|$ and so that we also have

$$\delta \sup_{v \in [\epsilon, W-\epsilon]} |\psi(v) - v^2 U'_-(v)| < \rho[\psi(W) - \psi(W - \epsilon)].$$
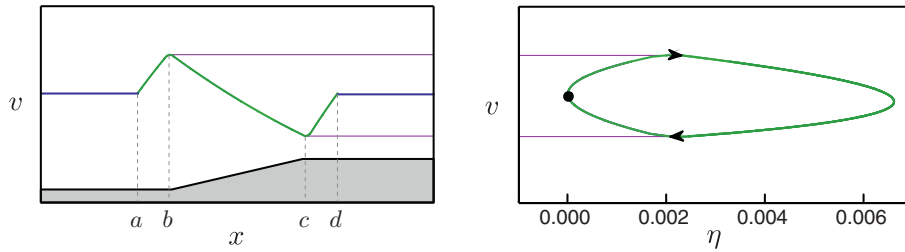
Fig. 9. Optimal speed profile on a steep uphill segment (left) and corresponding phase plot (right). The position x is measured in metres (m), the speed v is measured in metres per second (m s^{-1}) and the modified adjoint variable η is dimensionless.

Now (13) shows that $\zeta' \leq 0$ whenever $|\zeta| \leq \delta \Leftrightarrow -\delta \leq \zeta < 0$. It follows that $\zeta \leq -\delta$ for all $x > p$. Thus Maximum Brake must continue until $v = \epsilon$. Since $\epsilon > 0$ can be taken as small as we please it follows that a phase of Maximum Brake with $v < W$ will be the final phase and will terminate at $v = 0$.

2. A phase of Maximum Brake may terminate with $v \geq W$ by switching to Coast or HoldR. If $v > W$ at the termination point the switch must be to Coast.

3.12. Optimal switching points for phases of speedhold

We showed earlier that speedhold with positive control—where possible—is the most cost-efficient strategy whether or not the track is level. We have also seen that only two modes of singular speedhold control are permissible in a strategy of optimal type—a HoldP mode at the optimal driving speed $v = V$ using positive control $u = r(V) - g(x) > 0$ and a HoldR mode at the optimal regenerative braking speed $v = W = \psi^{-1}[\psi(V)/\rho] > V$ using negative control $u = r(W) - g(x) < 0$.

The HoldP mode is only possible if the track is not steep[8] at speed $v = V$. For segments of track that are steep at speed $v = V$ it is necessary to interrupt a HoldP phase by changing to a regular phase of Maximum Power with $u = U_+(v) > 0$ to traverse a steep uphill segment or to a regular phase of Coast with $u = 0$ to negotiate a steep downhill segment. On extended segments that are steep downhill at speed $v = V$ it is possible that the speed will rise during a regular Coast phase to the optimal regenerative braking speed $v = W$. If so it will be necessary to seek a strategy that transfers from Coast to HoldR at speed $v = W$.

The HoldR mode can only be used on track that is steep downhill[9] at speed $v = W$. For segments of track that are not steep at speed $v = W$ it is necessary to interrupt a HoldR phase by changing to a regular Coast phase with $u = 0$. On extended segments that are not steep at speed $v = W$ it is possible that the speed will decrease to the optimal driving speed $v = V$. If so it will be necessary to seek a strategy which transfers from Coast to HoldP at speed $v = V$.

Our main task is to find optimal switching points in two important situations—firstly when singular control phases are interrupted by regular phases of Maximum Power or Coast and secondly for transfers between singular control phases using a regular Coast phase. Our discussion in Part 1 of this paper will be largely intuitive and will focus on finding suitable intervals in which to search for the optimal switching points. The discussion is by no means complete—there are other more sophisticated methods that can be used to determine optimal switching points. We refer specifically to an important integral condition for optimality and to a closely-related intrinsic local energy minimization principle. Both the integral condition and the local energy minimization principle will be discussed in Part 2 of this paper (Albrecht et al., 2014).

#### 3.12.1. Interruptions to HoldP for a steep uphill segment

Consider an extended phase of HoldP at the optimal driving speed $v = V$ on non-steep track with $u = r(V) - g(x) > 0$. Assume we encounter a steep uphill segment of track (b, c). The track is steep uphill at speed $v = V$ on (b, c) if the net power is negative for all $x \in (b, c)$ with $u = U_+(V)$. In algebraic terms this means $U_+(V) - r(V) + g(x) < 0$. In physical terms it means we have insufficient power to maintain speeds $v \geq V$ at any point $x \in (b, c)$ using positive control. Hence HoldP is not possible on (b, c).

To traverse (b, c) we temporarily switch control from HoldP to Maximum Power with $u = U_+(v)$ on some interval $(a, d) \supset [b, c]$. During this phase the speed will rise above the optimal driving speed before entering the steep segment, fall below the optimal driving speed while traversing the steep segment, and then rise again after leaving the steep segment to return to the optimal driving speed. Thus we will have $v(a) = V$, $v(b) > V$, $v(c) < V$ and $v(d) = V$. Fig. 9 shows a stylised optimal speed profile and projected phase plot for the corresponding closed evolutionary line $(v(x), \eta(x))$ when HoldP is interrupted by Maximum Power to traverse a steep uphill segment.

A significant problem is to find the optimal switching points. There are many feasible intervals $(a, d) \supset [b, c]$ which satisfy the speed requirements. We can switch to Maximum Power at any point $a < b$ and as long as $v(c) < V$ we simply switch back to HoldP at the point $d > c$ where $v(d) = V$. However such a strategy is not necessarily optimal. The necessary conditions for optimal switching are defined by the evolution of both $v = v(x)$ and $\eta = \eta(x)$. We must choose (a, d) so that $v(a) = v(d) = V$

---

[8] The track is neither steep downhill nor steep uphill at speed $v = V$ if we have $-r(V) + g(x) \leq 0 \leq U_+(V) - r(V) + g(x)$.

[9] The track is steep downhill at speed $v = W$ if $U_-(W) - r(W) + g(x) < 0 < -r(W) + g(x)$. Note that we have assumed that the left-hand inequality is true everywhere.
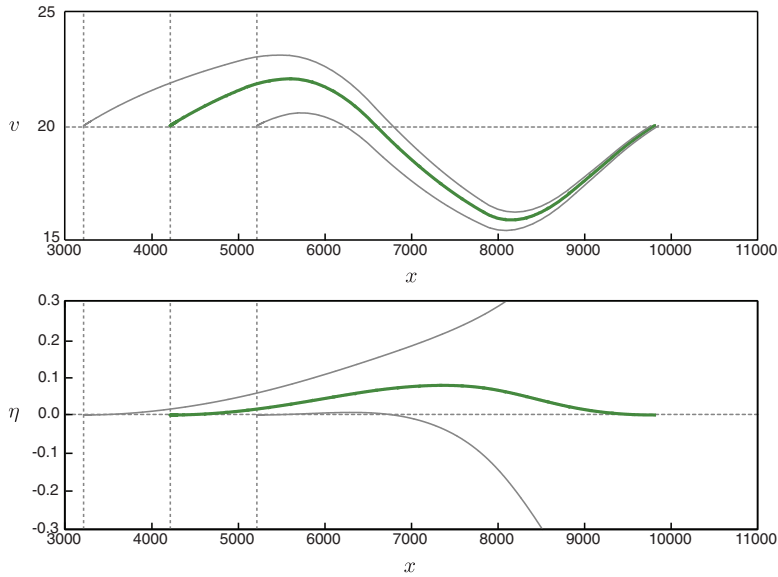
Fig. 10. Speed profiles $v = v(x)$ (top) and corresponding adjoint profiles $\eta = \eta(x)$ (bottom) for proposed phases of Maximum Power to traverse a steep uphill segment showing various starting points—a little too early (top curves), optimal (intermediate curves) and a little too late (bottom curves). The position x is measured in metres (m), the speed v is measured in metres per second (m s^{-1}) and the modified adjoint variable η is dimensionless.

and $\eta(a) = \eta(d) = 0$ with $\eta(x) > 0$ for all $x \in (a, d)$. We can see from (12) that we must also have $\eta'(a) = \eta'(d) = 0$. Thus the necessary conditions for optimal switching at $x = a$ and $x = d$ are also necessary for a minimum turning point of $\eta = \eta(x)$.

Fig. 10 shows three feasible speed profiles and the corresponding adjoint profiles for a phase of Maximum Power to traverse a steep uphill segment. Maximum Power starts a little too early on the top profile $(v_+(x), \eta_+(x))$ and a little too late on the bottom profile $(v_-(x), \eta_-(x))$. The intermediate profile $(v(x), \eta(x))$ is optimal because it terminates with $(v(d), \eta(d)) = (V, 0)$ as required. For the top profile $\eta_+(d_+) > 0$ when the speed returns to $v_+(d_+) = V$ and for the bottom profile $\eta_-(d_-) < 0$ when the speed returns to $v_-(d_-) = V$. Note how the speed profiles converge as x increases while the adjoint profiles diverge. This is a graphic demonstration that the state equations are stable in the forward direction whereas the adjoint equations are stable in the backward direction. We will extend our considerations by looking at earlier and later starting points. Our aim here is to extend the search region for the desired optimal strategy by finding reasonable upper and lower bounds on the optimal switching points.

Consider a Maximum Power phase with speed profile $v_- = v_-(x)$ that starts at the point $a_- = b$ and finishes at some point $d_- > c$ with $v = v_-(x) < V$ for all $x \in (a_-, d_-)$ and $v_-(a_-) = v_-(d_-) = V$. We will show that this starting point is too late. Since $\eta_-(a_-) = 0$ we can use (15) to deduce that

$$I_{a_-,+}(d_-)\eta_-(d_-) = \int_{a_-}^{d_-} \frac{\psi[v_-(\xi)] - \psi(V)}{v_-(\xi)^3} I_{a_-,+}(\xi) d\xi \tag{36}$$

where we have written $I_{a_-,+}(\xi)$ to denote the integrating factor $I_{p,u}(\xi)$ from (14) in the case where $p = a_-$ and $u(\xi) = U_+(v_-(\xi))$. Since $v_-(\xi) < V$ for all $\xi \in (a_-, d_-)$ it follows from (36) that $\eta_-(d_-) < 0$. Hence this strategy is not optimal.

Now suppose there is a Maximum Power phase with speed profile $v = v_+(x)$ that starts at some point $a_+ < b$ and finishes at the point $d_+ = c$ with $v = v_+(x) > V$ for all $x \in (a_+, d_+)$ and $v_+(a_+) = v_+(d_+) = V$. We will show that this starting point is too early.[10] Since $\eta_+(a_+) = 0$ we can use (15) to deduce that

$$I_{a_+,+}(d_+)\eta_+(d_+) = \int_{a_+}^{d_+} \frac{\psi[v_+(\xi)] - \psi(V)}{v_+(\xi)^3} I_{a_+,+}(\xi) d\xi \tag{37}$$

where we have used the notation $I_{a_+,+}(\xi)$ to denote the integrating factor $I_{p,u}(\xi)$ from (14) in the case where $p = a_+$ and $u(\xi) = U_+(v_+(\xi))$. Since $v_+(\xi) > V$ for all $\xi \in (a_+, d_+)$ it follows from (37) that $\eta_+(d_+) > 0$. Hence this strategy is not optimal.

Because the final value of η depends continuously on the starting position and because $\eta_-(d_-) < 0 < \eta_+(d_+)$ it follows that we can find a Maximum Power phase with speed profile $v = v(x)$ on an interval $(a, d) \supset [b, c]$ with $a \in (a_+, a_-)$ and $d \in (d_+, d_-)$ such that $(v(a), \eta(a)) = (v(d), \eta(d)) = (V, 0)$. This is the optimal phase. Stylised upper and lower bounding speeds are shown in Fig. 11.

---

[10] Although this strategy is an apparently suitable choice for an upper bound on the optimal speed profile it may not be possible to find a strategy with $v_+(c) = V$ if the hill is too long or too steep. An alternative upper bound will be introduced in Part 2 of this paper (Albrecht et al., 2014).
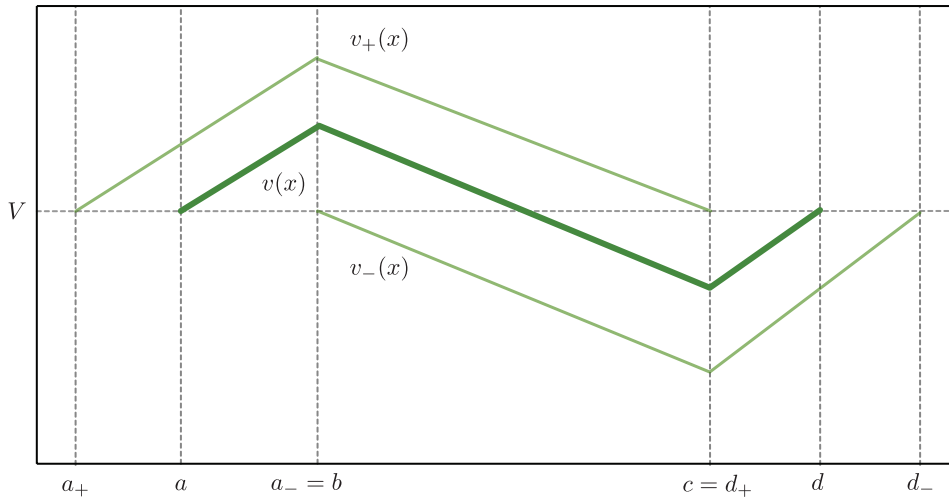
Fig. 11. Stylised bounding speed profiles $v = v_-(x)$ on the interval $[a_-, d_-]$ where $a_- = b$ and $v = v_+(x)$ on the interval $[a_+, d_+]$ where $d_+ = c$ for a Maximum Power phase over a steep uphill segment (b, c). The stylised optimal speed profile $v = v(x)$ is also shown on the interval $[a, d]$. The position x is measured in metres (m) and the speed v is measured in metres per second (m s^{-1}).
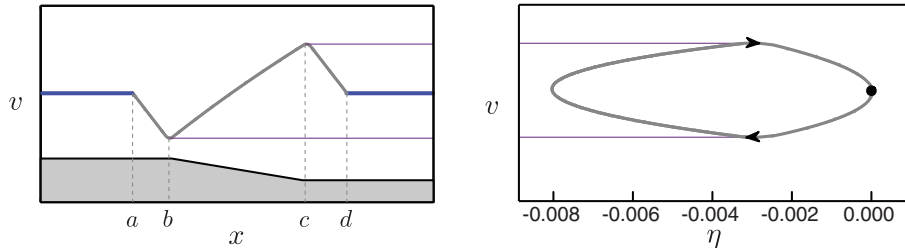
Fig. 12. Optimal speed profile on a steep downhill segment (left) and corresponding phase plot (right). The position x is measured in metres (m), the speed v is measured in metres per second (m s^{-1}) and the modified adjoint variable η is dimensionless.

3.12.2. Interruptions to HoldP for a steep downhill segment

Consider an extended phase of HoldP on non-steep track. Assume we encounter a steep downhill segment of track (b, c). The track is steep downhill at speed $v = V$ on (b, c) if the net power is positive for all $x \in (b, c)$ with $u = 0$. In algebraic terms this means that $-r(V) + g(x) > 0$. In physical terms it means that it is not possible to maintain any speed $v \leq V$ at point x with non-negative control. Hence HoldP is not possible on (b, c).

To traverse (b, c) we temporarily switch control from HoldP to Coast with $u = 0$ on some interval $(a, d) \supset [b, c]$. During this phase the speed will fall below the optimal driving speed before entering the steep segment (b, c), rise above the optimal driving speed while traversing (b, c), and fall again after leaving (b, c) to return to the optimal driving speed. Thus $v(a) = V$, $v(b) < V$, $v(c) > V$ and $v(d) = V$. There are many feasible intervals $(a, d) \supset [b, c]$ which will satisfy the speed requirements. We can switch to Coast at any point $a < b$ and as long as $v(c) > V$ we then switch back to HoldP at the point $d > c$ where $v(d) = V$. However such a strategy is not necessarily optimal. Fig. 12 shows a stylised optimal speed profile and corresponding phase plot $(v(x), \eta(x))$ for the closed evolutionary line when HoldP is interrupted by Coast to traverse a steep downhill segment.

We need to find the optimal switching points. The necessary conditions for optimal switching are defined by $\eta(a) = \eta(d) = 0$ with $\rho - 1 < \eta(x) < 0$ for all $x \in (a, d)$. We can see from (12) that $\eta'(a) = \eta'(d) = 0$. These conditions are also necessary for a minimum of $\eta = \eta(x)$ at $x = a$ and $x = d$. We will assume that $\eta(x) > \rho - 1 \Leftrightarrow \zeta > 0$ for all $x \in (b, c)$ during the inserted Coast phase. If there is some $x \in (b, c)$ where $\eta(x) = \rho - 1 \Leftrightarrow \zeta = 0$ then we will need to consider an alternative strategy.

On track with piecewise-constant gradient the evolutionary lines can be used to find optimal switching points, albeit from a less general point of view. Previously, for a steep uphill segment, we proposed a nominal starting point and followed the speed profile $v = v(x)$ for the inserted Maximum Power phase until the speed returned to the optimal hold speed $v = V$. Then we found a corresponding profile for the modified adjoint variable $\eta = \eta(x)$ and checked to see if it started and finished at the critical value $\eta = 0$. If not we chose another starting point and repeated the above procedure until we found a starting point such that $v = v(x)$ and $\eta = \eta(x)$ did start and finish at the desired values. A similar procedure could be used here to find the optimal Coast phase. However this time, for a steep downhill segment on track with piecewise-constant gradient, we will simply follow the evolutionary line $(v(x), \eta(v(x)))$ for a proposed Coast phase until either we return to the critical point $(v, \eta) = (V, 0)$ as required for optimality or until we can see that this will not happen.
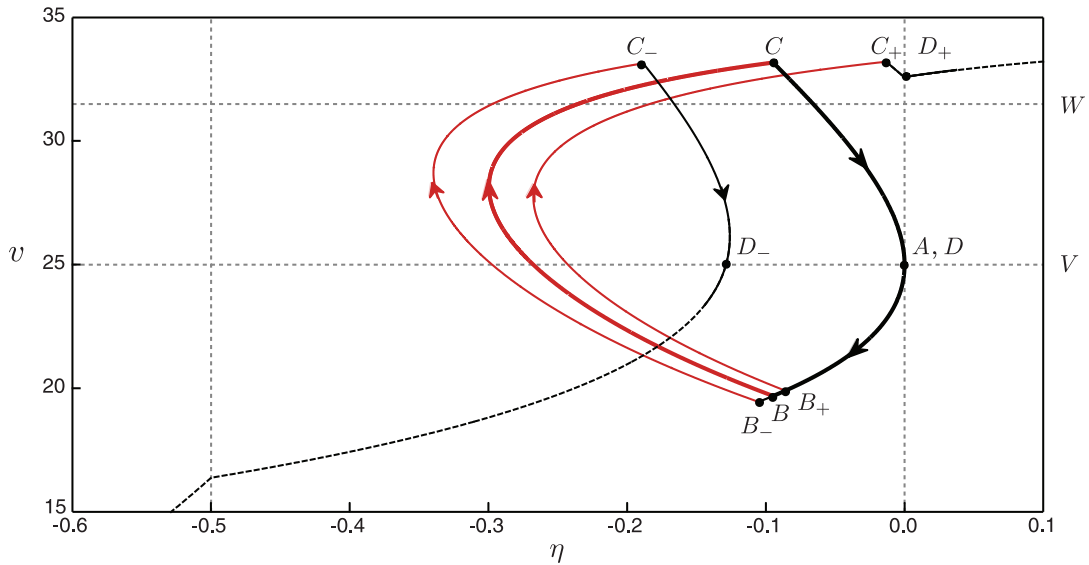
Fig. 13. Phase plots for three possible evolutionary lines where HoldP is interrupted by Coast to traverse a steep downhill segment. The closed curve *ABCD* is optimal. The open curves $AB_-C_-D_-$ and $AB_+C_+D_+$ start Coast too early and too late respectively. Note that the starting point *A* on the phase plot represents three different physical points—one with coordinates $(x, v, \eta) = (a_-, V, 0)$, one with coordinates $(x, v, \eta) = (a, V, 0)$ and one with coordinates $(x, v, \eta) = (a_+, V, 0)$ where $a_- < a < a_+$. Note that the curves $AB_-C_-D_-$ and $AB_+C_+D_+$ are shown as dashed lines beyond their nominal termination points to show that they can never return to the desired termination point *A*. The modified adjoint variable η is dimensionless and the speed *v* is measured in metres per second (m s^{-1}).

For the sake of simplicity, suppose the track is level except for a steep downhill segment (b, c) with constant gradient acceleration $g(x) = \gamma > 0$. Fig. 13 shows three possible evolutionary lines for Coast phases starting with phase coordinates $(v, \eta(v)) = (V, 0)$ at the critical point *A*. Although these lines start with the same phase coordinates we emphasise that the switching points on the track are different. The first evolutionary line is for a phase that starts too early at $x = a_-$ with $(v, \eta) = (V, 0)$; the switching point $x = a_-$ is too far from the steep segment. The second evolutionary line is for a phase that starts too late at $x = a_+$ with $(v, \eta) = (V, 0)$; the switching point $x = a_+$ is too close to the steep segment. The third is optimal and starts at $x = a \in (a_-, a_+)$ with $(v, \eta) = (V, 0)$; the only line that returns to the critical point *A* with coordinates $(v, \eta) = (V, 0)$. The phase coordinates are denoted by (v_-, η_-) for the line that starts too early at $x = a_-$, by (v_+, η_+) for the line that starts too late at $x = a_+$ and by (v, η) for the optimal line[11] starting at $x = a \in (a_-, a_+)$. We remind readers that track positions—the *x* coordinates that generate the phase points—are not shown on the phase plot.

The optimal Coast phase with coordinates (v, η) begins at some switching point $x = a < b$ with $(v(a), \eta(v(a))) = (V, 0)$. The evolutionary coordinates (v, η) follow a segment *AB* along an evolutionary line for Coast on level track starting at the critical point *A* with $x = a$ and finishing at a point *B* at the start of the steep downhill segment. On this segment both $v = v(x)$ and $\eta = \eta(v(x))$ will decrease. The point *B* at $x = b$ has coordinates $(v, \eta) = (v(b), \eta(v(b)))$ calculated by solving

$$b = a + \int_v^V \frac{wdw}{r(w)}$$

for $v = v(b) < V$ and then using the formula

$$\eta(v(b)) = \frac{E_V(v(b)) - E_V(V)}{-r(v(b))}.$$

From the point *B* the coordinates (v, η) follow a segment *BC* along an evolutionary line for Coast on steep downhill track with constant gradient $g(x) = \gamma > 0$ to a point *C* at the end of the steep downhill segment. On this segment $v = v(x)$ increases while $\eta = \eta(v)$ initially decreases but then increases. The point *C* at $x = c$ has coordinates $(v, \eta) = (v(c), \eta(c))$ calculated by solving

$$c = b + \int_{v(b)}^v \frac{wdw}{-r(w) + \gamma}$$

for $v = v(c) > V$ and then using the formula

$$\eta(v(c)) = \frac{E_V(v(c)) - E_V(V) + \gamma \eta(v(b))}{-r(v(c)) + \gamma}.$$

Note that the segment (b, c) is steep downhill at speed $v = V$. Since $v(b) < V$ it follows that $-r(w) + \gamma > 0$ for all $w \in [v(b), v(c)] \subset (0, V_{0, \gamma})$ where $V_{0, \gamma} > V$ is the limiting speed on this line. Finally, from the point *C*, the coordinates (v, η) follow a segment CD along

---

[11] We use the notation $v_- = v_-(x)$ and $v_+ = v_+(x)$ because it turns out that $v_-(x) \leq v(x) \leq v_+(x)$ for all *x* where $v = v(x)$ is the optimal profile.

the evolutionary line for Coast on level track that terminates at the point D with $x = d$ and coordinates $(v, \eta) = (v(d), \eta(d)) = (V, 0)$. This is a segment of the same evolutionary line on which the optimal Coast phase began. In the phase plane the critical point D coincides with the critical point A.

The Coast phase with coordinates denoted by (v_-, η_-) starts at some switching point $x = a_- < a$ with $(v_-(a_-), \eta_-(v_-(a_-))) = (V, 0)$ and follows a segment AB_- on the same evolutionary line as the optimal trajectory to a point B_- with $x = b$ at the start of the steep segment. On this segment both $v_- = v_-(x)$ and $\eta_- = \eta_-(v_-)$ decrease. Because we started Coast further away from $x = b$ we will have $v_-(x) < v(x)$ for each $x \in [a, b]$ and in particular we will have $v_-(b) < v(b)$ at the point B_-. The adjoint variable will fall to a value $\eta_-(v_-(b)) < \eta(v(b))$. The coordinates $(v_-(b), \eta_-(v_-(b)))$ are calculated in the manner described above for the corresponding point on the optimal line. From this point the coordinates (v_-, η_-) follow a segment B_-C_- along an evolutionary line for Coast on the steep downhill track to a point C_- with $x = c$ at the end of the steep segment. On this segment $v_- = v_-(x)$ increases while $\eta_- = \eta_-(v_-)$ initially decreases but then increases. The point C_- has coordinates $(v_-(c), \eta_-(v_-(c)))$ calculated in a similar way to the way we calculated the coordinates for C. Once again, the limiting speed on this evolutionary line is $V_{0,\gamma} > V$. Note that $v_-(c) < v(c)$ and $\eta_-(v_-(c)) < \eta(v(c)) < 0$. From the point C_-, the coordinates (v_-, η_-) follow a segment C_-D_- along an evolutionary line for a Coast curve on level track that we nominally terminate at some point D_- with $x = d_-$ and coordinates $(v, \eta) = (v_-(d_-), \eta_-(d_-)) = (V, \eta_-(d_-))$ where $\eta_-(d_-) < 0$.

Remark 6. In Fig. 13 the Coast curve B_-C_- on the steep downhill track for $x \in [b, c]$ lies everywhere to the left of the optimal curve BC and finishes with $(v, \eta) = (v_-(c), \eta_-(c))$ where $v_-(c) < v(c)$ and $\eta_-(c) < 0$. Since $\eta_-(c) < 0$ we transfer to a Coast curve C_-D_- which lies everywhere to the left of the optimal curve CD. This curve eventually reaches some point D_- with $x = d_-$ and coordinates $(v, \eta) = (v_-(d_-), \eta_-(d_-))$ where $v_-(d_-) = V$ and $\eta_-(d_-) < 0$. Since $\eta_-(d_-) < 0$ the evolutionary line will continue beyond $x = d_-$ on the same Coast curve with both v and η continuing to decrease. Hence we cannot return to the point $(V, 0)$ and so we terminate the curve at D_-. Since $v_-(c) < v(c)$ it follows that

$$d_- = c + \int_V^{v_-(c)} \frac{dw}{r(w)} < c + \int_V^{v(c)} \frac{w dw}{r(w)} = d.$$

In Fig. 13 the continuation of the Coast curve C_-D_- beyond the nominal termination point D_- is shown as a dashed line.

The Coast phase with coordinates denoted by (v_+, η_+) starts at a switching point $x = a_+ \in (a, b)$ with $(v_+(a), \eta_+(v_+(a))) = (V, 0)$. The evolutionary coordinates (v_+, η_+) follow a segment AB_+ along the same evolutionary line as the optimal trajectory to a point B_+ with $x = b$ at the start of the steep segment. On this segment both $v_+ = v_+(x)$ and $\eta_+ = \eta_+(v_+)$ will decrease. Because we started Coast closer to $x = b$ we will have $v_+(x) > v(x)$ for each $x \in [a_+, b]$ and in particular $v_+(b) > v(b)$ at the point B_+. The adjoint variable will fall to a value $\eta_+(v_+(b)) > \eta(v(b))$. The coordinates $(v_+(b), \eta_+(v_+(b)))$ are calculated in the manner described above for the corresponding points on the optimal curve. From this point the coordinates (v_+, η_+) follow a segment B_+C_+ along an evolutionary line for Coast on the steep downhill track to a point C_+ with $x = c$ at the end of the steep segment. On this segment $v_+ = v_+(x)$ increases while $\eta_+ = \eta_+(v_+)$ initially decreases but then increases. The point C_+ has coordinates $(v_+(c), \eta_+(v_+(c)))$ calculated in the same way as we calculated the corresponding coordinate points on the optimal curve. As before the limiting speed on this evolutionary line is $V_{0,\gamma} > V$. Note that $v_+(c) > v(c)$ and $\eta_+(v_+(c)) > \eta(v(c))$. If we are sufficiently close to the optimal curve we will still have $\eta_+(v_+(c)) < 0$. Thus, (v_+, η_+) now follows a segment C_+D_+ along an evolutionary line on level track that we nominally terminate at some point D_+ with $x = d_+$ and coordinates $(v(d_+), \eta_+(d_+)) = (v_+(d_+), 0)$ where $v_+(d_+) > V$.

Remark 7. In Fig. 13 the Coast curve B_+C_+ on the steep downhill track for $x \in [b, c]$ lies everywhere to the right of the optimal curve BC and finishes with $(v, \eta) = (v_+(c), \eta_+(c))$ where $v_+(c) > v(c)$ and (provided the curve is sufficiently close to the optimal curve BC) with $\eta_+(c) < 0$. Since $\eta_+(c) < 0$ we transfer to a Coast curve C_+D_+ which lies everywhere to the right of the optimal curve CD. This curve eventually reaches some point with $x = d_+$ and coordinates $(v, \eta) = (v_+(d_+), \eta_+(d_+))$ where $v_+(d_+) = V + h > V$ and $\eta_+(d_+) = 0$. At this point the Coast curve on level track transforms into a Maximum Power curve on level track and so both v and η will increase. Hence this curve can never return to the point $(V, 0)$ and so we terminate the curve at D_+. The position on the track is defined by

$$d_+ = c + \int_{V+h}^{v_+(c)} \frac{dw}{r(w)}.$$

While the position $x = d_+$ is not entirely clear in relation to the point $x = d$ where the optimal curve finishes it seems intuitively reasonable from Fig. 13 that if the curve C_+D_+ is well to the right of CD with $\eta_+(d_+) \approx 0$ then we will have $V + h \approx v_+(c)$ and

$$d_+ = c + \int_{V+h}^{v_+(c)} \frac{dw}{r(w)} = c + \delta \approx c.$$

If the curve C_+D_+ is close to CD then we will have $v_+(c) \approx v(c)$ and $V + h \approx V$ so

$$d_+ = c + \int_{V+h}^{v_+(c)} \frac{dw}{r(w)} \approx c + \int_V^{v(c)} \frac{dw}{r(w)} = d.$$

In Fig. 13 the continuation of the Coast curve C_+D_+ as a Maximum Power curve for $x > d_+$ beyond the nominal termination point D_+ is shown as a dashed line.

Once we have found Coast phases that start too early and too late we can converge to the optimal Coast phase using an elementary midpoint iteration on the starting point.

3.12.3. Interruptions to HoldR for a non-steep segment

Let us assume that during an extended phase of HoldR at the optimal regenerative braking speed $v = W$ with $u = r(W) - g(x) < 0$ we encounter a segment of track (b, c) that is not steep downhill at speed $v = W$.

To traverse the non-steep segment (b, c) it is necessary to switch control from HoldR to Coast with $u = 0$ on some interval $(a, d) \supset [b, c]$. During this phase the speed will rise above the optimal regenerative braking speed before reaching the non-steep segment (b, c), fall below the optimal regenerative braking speed on (b, c), and then rise again after leaving (b, c) to return to the optimal regenerative braking speed. Thus $v(a) = W$, $v(b) > W$, $v(c) < W$ and $v(d) = W$. There are many Coast phases on intervals $(a, d) \supset [b, c]$ which satisfy the speed requirements. We can switch to Coast at any point $a < b$ and as long as $v(c) < W$ we simply switch back to HoldR at the point $d > c$ where $v(d) = W$.

Once again we need to find the optimal switching points. The necessary conditions for optimal switching are $v(a) = v(d) = W$ and $\zeta(a) = \zeta(d) = 0$ with $0 < \zeta < 1 - \rho$ for all $x \in (a, d)$. We can see from (13) that $\zeta'(a) = \zeta'(d) = 0$. Thus the necessary conditions for optimal switching are also necessary for a minimum of $\zeta = \zeta(x)$. It can be seen that a Coast phase with initial switching point $a_- = b$ will have $v_-(a_-) = v_-(d_-) = W$ for some $d_- > c$ and $v_-(x) < W$ for all $x \in (a_-, d_-)$. Hence we can use (16) and an argument similar to that used earlier to show that $\zeta_-(d_-) < 0$. It follows that this phase starts too late. If we can find a Coast phase with final switching point $d_+ = c$ we will have $v_+(a_+) = v_+(d_+) = W$ for some $a_+ < b$ and $v_+(x) > W$ for all $x \in (a_+, d_+)$. Therefore we can use (16) and an argument similar to that used earlier to show that $\zeta_+(d_+) > 0$. Thus we deduce that this phase starts too early.[12] Hence there is an optimal Coast phase with an initial switching point $a \in (a_+, a_-)$ and a final switching point $d \in (d_+, d_-)$.

3.12.4. Changing from HoldP to HoldR

Consider a train following a strategy of optimal type in a HoldP phase at the optimal driving speed $v = V$ on non-steep track that enters a segment (b, c) which is not only steep downhill at speed $v = V$ but also contains an extended segment which is steep downhill at the optimal regenerative braking speed $v = W > V$. We wish to find an optimal Coast phase that allows a transition from HoldP at speed $v = V$ to HoldR at speed $v = W$.

Suppose there is a Coast phase $(v_+(x), \eta_+(x))$ with initial switching point at $p_+ = b$ and with $(v_+(b), \eta_+(b)) = (V, 0)$. We will show that this phase starts too late and finishes too early. The track is steep downhill at speed $v = V$ for $x > p_+ = b$ and so $v_+(x) > V$ for $x > p_+$. Because the train enters an extended segment of track which is steep downhill at speed $v = W$ we assume it eventually reaches a point $x = q_+ \in (b, c)$ where $v(q_+) = W$. By using the integrating factor $I_{p_+, 0}(x)$ from (14) to solve (12) with $u(v) = 0$ we obtain

$$I_{p_+, 0}(x)\eta_+(x) = \int_{p_+}^{x} \frac{\psi[(v_+(\xi)] - \psi(V)}{v_+(\xi)^3} I_{p_+, 0}(\xi) d\xi \tag{38}$$

for all $x \in [p_+, q_+]$. Since $v_+(x) > V$ for all $x \in (p_+, q_+) \subset (b, c)$ and because $I_{p_+, 0}(x) > 0$ it follows that $\eta_+(q_+) > 0 > \rho - 1$.

To complete the argument we must assume there is a Coast phase starting too early[13] at some switching point $p_- < b$ with $(v_-(p_-), \eta_-(p_-)) = (V, 0)$ and finishing at some point $q_- \in (b, c)$ with $v_-(q_-) = W$ and $\eta_-(q_-) < \rho - 1$. The earlier start means that the speed decreases initially so that $v_-(b) < V = v_+(b)$. Therefore the speed profile $v = v_-(x)$ is always below the speed profile $v = v_+(x)$ and hence it finishes later at some point $q_- > q_+$ with $v_-(q_-) = W$. It follows that there is a Coast phase starting at some initial optimal switching point $p \in (p_-, p_+)$ with $(v(p), \eta(p)) = (V, 0)$ and finishing at a final switching point $q \in (q_+, q_-)$ with $(v(q), \eta(q)) = (W, \rho - 1)$. This phase allows a strategy of optimal type to transfer from HoldP at the optimal driving speed to HoldR at the optimal regenerative braking speed. See Fig. 1 for an optimal strategy that includes an optimal transfer HoldP–Coast–HoldR.

3.12.5. Changing from HoldR to HoldP

Consider a train following a strategy of optimal type in a HoldR phase at the optimal regenerative braking speed $v = W$ with $u = r(W) - g(x) < 0$ on steep downhill track and suppose the train approaches a segment (b, c) which is not steep downhill at the optimal regenerative braking speed $v = W$ and which also contains an extended segment that is not steep downhill at the lower optimal driving speed $v = V$.

Suppose there is a Coast phase $(v_-(x), \zeta_-(x))$ that starts at the switching point $p_- = b$ with $(v_-(b), \zeta_-(b)) = (W, 0)$. We will show that this phase starts too late and finishes too early. The track is not steep downhill at speed $v = W$ for $x > b$ and so $v_- = v_-(x)$ must initially decrease with $v_-(x) < W$ for all $x \in (b, c)$. Because the train moves onto an extended segment of track which is no longer steep downhill at speed $v = V$ we assume it eventually reaches a point $q_- \in (b, c)$ with $v_-(q_-) = V$. Since $\zeta_-(p_-) = 0$ we can use (16) to deduce that

$$I_{p_-, 0}(x)\zeta_-(x) = \rho \int_{p_-}^{x} \frac{\psi[(v_-(\xi)] - \psi(W)}{v_-(\xi)^3} I_{p_-, 0}(\xi) d\xi \tag{39}$$

---

[12] Although this strategy provides a convenient upper bound to the optimal speed profile it may not be possible to find a Coast strategy with $v_+(c) = W$ if the non-steep section is too long. An alternative upper bound will be introduced in Part 2 of this paper (Albrecht et al., 2014).

[13] The existence of a Coast strategy that starts too early is far from obvious. A more comprehensive argument will be proposed in Part 2 of this paper (Albrecht et al., 2014).

for all $x \in (p_-, q_-)$ where we have used the notation $I_{p_-,0}(\xi)$ to denote the integrating factor $I_{p,u}(\xi)$ from (14) in the case where $p = p_-$ and $u = 0$. Since $v_-(x) < W$ for all $x \in (p_-, q_-) \subset (b, c)$ and since $I_{p_-,0}(x) > 0$ it follows that $\zeta_-(q_-) < 0 < 1 - \rho$.

We must also assume there is a Coast phase starting at some switching point $p_+ < b$ with $(v_+(p_+), \zeta_+(p_+)) = (W, 0)$ and $v_+(b) > W$ which finishes at a point $q_+ \in (b, c)$ with $v_+(q_+) = V$ and $\zeta_+(q_+) > 1 - \rho$. The earlier start means that the speed increases initially so that $v_+(b) > W = v_-(b)$. Therefore the speed profile $v = v_+(x)$ is always above the speed profile $v = v_-(x)$ and hence it finishes later at some point $x = q_+ > q_-$ with $v_+(q_+) = V$. This speed profile provides a convenient upper bound[14] for the optimal speed profile. It follows that there is a Coast phase starting at some initial switching point $p \in (p_+, p_-)$ with $(v(p), \zeta(p)) = (W, 0)$ and finishing at a final switching point $q \in (q_-, q_+)$ with $(v(q), \zeta(q)) = (V, 1 - \rho)$. This phase allows a strategy of optimal type to transfer from HoldR at the optimal regenerative braking speed to HoldP at the optimal driving speed. See Fig. 1 for an optimal strategy that includes an optimal transfer HoldR–Coast–HoldP.

3.13. Speed limits

If we impose a maximum-speed constraint $v(x) \leq v_m(x)$ where $v_m(x)$ is a known function[15] then it follows from Girsanov et al. (1972), Hartl et al. (1995) and Khmelnitsky (2000) that the Hamiltonian $\mathcal{H}$ remains the same as in (5) but the Lagrangian takes the modified form

$$\mathcal{L} = \mathcal{H} + \pi_1(U_+(v) - u) + \pi_2(u - U_-(v)) + \sigma(v_m - v) \tag{40}$$

where $\pi_1 \geq 0$, $\pi_2 \geq 0$ and $\sigma \geq 0$ are Lagrange multipliers and the adjoint variables $(\mu_1, \mu_2) = (\mu_1(x), \mu_2(x))$ now satisfy the differential equations

$$\mu_1' = (-1)\frac{\partial \mathcal{L}}{\partial t} = 0 \tag{41}$$

$$\mu_2' = (-1)\frac{\partial \mathcal{L}}{\partial v} = \frac{\mu_1}{v^2} + \frac{\mu_2(u - r(v) + g(x))}{v^2} + \frac{\mu_2 r'(v)}{v} - \pi_1 U_+'(v) + \pi_2 U_-'(v) + \sigma \tag{42}$$

except possibly for finite jumps at entry and exit points for intervals where the speed constraint is active. In order to maximise the Hamiltonian subject to the necessary control constraints and the imposed speed limit we see that the necessary condition

$$\frac{\partial \mathcal{L}}{\partial u} = (-1)\left[\frac{(1 + \text{sgn}(u))}{2} + \frac{\rho(1 - \text{sgn}(u))}{2}\right] + \frac{\mu_2}{v} - \pi_1 + \pi_2 = 0 \tag{43}$$

remains unchanged from (9). Thus we have $-1 + \mu_2/v - \pi_1 + \pi_2 = 0$ when $u > 0$ and $-\rho + \mu_2/v - \pi_1 + \pi_2 = 0$ when $u < 0$. Hence we obtain the same five basic optimality conditions.

3.13.1. Control phases for an optimal strategy with speed limits

Despite the speed limit constraint the analysis remains the same for optimal phases of regular control. Thus we have phases of Maximum Power with $\mu_2 > v \Leftrightarrow \eta > 0$ and $u = U_+(v) > 0$; Coast with $v > \mu_2 > \rho v \Leftrightarrow 0 > \eta > \rho - 1 \Leftrightarrow 1 - \rho > \zeta > 0$ and $u = 0$; and Maximum Brake with $\rho v > \mu_2 \Leftrightarrow 0 > \zeta$ and $u = U_-(v) < 0$. For each of these phases we must have $v < v_m$ and $\sigma = 0$ except possibly at the endpoints.

For optimal phases of singular control there are significant additional considerations. On intervals where $v < v_m$ the Karush–Kuhn–Tucker conditions give $\sigma(v_m - v) = 0 \Rightarrow \sigma = 0$ and so the optimal strategy remains essentially the same. Thus for $\eta = 0$ we have an optimal phase of HoldP with $v = V$, $\mu_1 = -\psi(V)$ and $0 < u = r(V) - g(x) < U_+(V)$ and for $\eta = \rho - 1 \Leftrightarrow \zeta = 0$ we have an optimal phase of HoldR with $v = W$, $\mu_1 = -\psi(V) = -\rho\psi(W)$ and $U_-(v) < u = r(W) - g(x) < 0$. On intervals of Constrained Speed where $v = v_m$ the control $u = u_m$ is defined by $u_m(x) = v_m(x)v_m'(x) + r[v_m(x)] - g(x)$. We must have $U_-(v) < u_m < U_+(v)$ and so $\pi_1 = \pi_2 = 0$.

If $u_m > 0$ and $v = v_m$ then $\partial \mathcal{L}/\partial u = 0 \Rightarrow \mu_2/v_m = 1 \Leftrightarrow \eta = 0$. Now (42) shows

$$\sigma = \frac{\psi(V)}{v_m^2} - r'(v_m) \iff \sigma = -E_V'(v_m).$$

Since $\sigma \geq 0$ and $E_V(v)$ is strictly quasiconvex with a unique minimum at $v = V$ it is necessary that $v_m \leq V$ so that $E_V'(v_m) < 0$.

If $u_m < 0$ then $\partial \mathcal{L}/\partial u = 0 \Leftrightarrow \mu_2/v_m = \rho \Leftrightarrow \zeta = 0$. In this case (42) shows

$$\sigma = \rho\left[\frac{\psi(W)}{v_m^2} - r'(v_m)\right] \iff \sigma = -\rho E_W'(v_m).$$

Since $\sigma \geq 0$ and $E_W(v)$ is strictly quasiconvex in v with a unique minimum at $v = W$ it is necessary that $v_m \leq W$ so that $E_W'(v_m) < 0$.

These results extend the corresponding results obtained by Khmelnitsky (2000). An important point to remember is that a Constrained Speed phase is simply another form of singular control—just as HoldP and HoldR are singular controls—where the

---

[14] A more comprehensive argument about the existence of a suitable upper bound will be given in Part 2 of this paper (Albrecht et al., 2014).

[15] We will assume that the function $v_m = v_m(x)$ is continuous and piecewise smooth on all relevant intervals. Thus, in particular, $u_m = u_m(x) = v_m(x)v_m'(x) + r[v_m(x)] - g(x)$ is a piecewise-continuous function of x. We assume $v_m(x) \leq \bar{v}(x)$ and $U_-(v_m) \leq u_m \leq U_+(v_m)$ for $v_m = v_m(x)$ when $x \in [0, X]$.

level of control is determined by the desired speed profile. We may choose to leave a Constrained Speed phase whenever we wish. Our decision is determined by the need to find a feasible strategy while continuing to follow a strategy of optimal type.

Remark 8. Suppose we have $v(x) \leq v_m(x)$ where $v_m(x)$ is an imposed speed limit. For each strategy of optimal type there is a uniquely-defined optimal driving speed $v = V$ that must be used on every HoldP phase. In cases where $V > v_m(x)$ for all $x \in [0, X]$ there will be no HoldP phases but changes to V may still cause changes to the optimal strategy. As V increases the journey time decreases. If there are no HoldP phases for the given value of V then a further increase in V simply means there will be less coasting. If $V > v_m(x)$ on some restricted interval $x \in (p, q)$ then the train may follow the speed limit with $v(x) = v_m(x)$ for $x \in [p, q]$ but can never exceed it. If the journey time is decreased then V will increase and all corresponding increases in actual speed will be at points where $v(x) < v_m(x)$.

3.13.2. Phase transitions for an optimal strategy with speed limits

Since many of the functions in this section are piecewise continuous and since we will need to consider possible jump discontinuities at points where the control changes it is convenient to use the notation $f(a - 0) = \lim_{h \downarrow 0} f(a - h)$ for the left-hand limit and $f(a + 0) = \lim_{h \downarrow 0} f(a + h)$ for the right-hand limit. Suppose an optimal phase of Maximum Power with $u = U_+(v)$ changes to a phase of Constrained Speed with $v = v_m < V$ and $u = u_m = v_m v_m' + r(v_m) - g(x) > 0$ at some point $x = a$ where $v_m(a) < V$. For $x < a$ we have $u(x) = U_+(v(x)) \Rightarrow u(a - 0) = U_+(v_m(a))$ and since we also have $\mu_2(x) > v(x)$ it follows that $\mu_2(a - 0) \geq v_m(a) \Leftrightarrow \eta(a - 0) \geq 0$. For $x > a$ we have $u = u_m > 0$ and so $u(a + 0) = u_m(a)$. We also know that $\mu_2 = v_m \Leftrightarrow \eta = 0$ for $x > a$ and so $\mu_2(a + 0) = v_m(a) \Leftrightarrow \eta(a + 0) = 0$. Therefore at $x = a$ there is a possible jump discontinuity in the Hamiltonian given by

$$
\begin{aligned}
\mathcal{H}(a + 0) - \mathcal{H}(a - 0) &= (-1)u(a + 0) - \frac{\psi(V)}{v_m(a)} + \frac{\mu_2(a + 0)[u_m(a) - r(v_m(a)) + g(a)]}{v_m(a)} \\
&\quad - \left\{ (-1)u(a - 0) - \frac{\psi(V)}{v_m(a)} + \frac{\mu_2(a - 0)[U_+(v_m(a)) - r(v_m(a)) + g(a)]}{v_m(a)} \right\} \\
&= (-1)[v_m(a)v_m'(a) + r(v_m(a)) - g(a)] + v_m(a)v_m'(a) \\
&\quad - \left\{ (-1)U_+(v_m(a)) + \frac{\mu_2(a - 0)[U_+(v_m(a)) - r(v_m(a)) + g(a)]}{v_m(a)} \right\} \\
&= (-1)\left[\frac{\mu_2(a - 0)}{v_m(a)} - 1 \right][U_+(v_m(a)) - r(v_m(a)) + g(a)] \\
&= [\eta(a + 0) - \eta(a - 0)][U_+(v_m(a)) - r(v_m(a)) + g(a)].
\end{aligned}
$$

Now suppose an optimal phase of Coast with $u = 0$ changes to a phase of Constrained Speed with $v = v_m$ at some point $x = a$ where $v_m(a) < W$. A similar argument gives

$$
\mathcal{H}(a + 0) - \mathcal{H}(a - 0) = [\zeta(a + 0) - \zeta(a - 0)][-r(v_m(a)) + g(a)].
$$

Other transitions to and from a Constrained Speed phase are analysed in a similar way.

4. Conclusions and future work

In this two-part paper our intention has been to review the key principles of optimal train control and extend the known results to a more general model. In so doing we were able to highlight the essential properties of the model.

In Part 1 we formulated a general model that captures most of the known characteristic behaviour and then derived necessary conditions for optimal control. We showed that each strategy of optimal type consists of a (possibly long) sequence of distinct phases constructed using five basic optimal control modes and that the precise structure of the strategy is determined by the joint evolution of the state and adjoint variables. We solved the state and adjoint equations and used the solutions and the corresponding evolutionary lines to find optimal switching points for the controls. In Part 2 we will prove that an optimal strategy always exists and develop alternative forms of the necessary conditions for optimal switching. In particular we will derive a local energy minimization principle and use it to show that the optimal strategy is uniquely defined.

We believe that there is value in a coherent summary of the work on optimal train control and that the timing is appropriate. Much of the modern theory has been developed over the past three decades and the results obtained by Howlett et al., Khmelnitsky, and Liu and Golovitcher concerning optimal driving strategies for a single train are now widely recognised as fundamental. To obtain the full benefits from our comprehensive review and extended discussion it is vital that the rail research community communicate these ideas and the principles of energy-efficient driving practice to train operators and track providers. The practical need for energy conservation and efficient operation has never been more urgent. The most pressing research challenges for the future in this area are to develop optimal control policies for trains travelling in the same direction on the same line in such a way that safe separation is maintained between trains. On busy rail networks, solution of the train separation problem relates to and depends on integrated scheduling and control policies to ensure that train movements are both energy-efficient and effectively coordinated.

Acknowledgements

References

Albrecht, A., Howlett, P., Pudney, P., Vu, X., Zhou, P., 2013. The two train separation problem on level track. In: Gaertner, P., Bowden, F., Piantadosi, J., Mobbs, K. (Eds.), 22nd National Conference of the Australian Society for Operations Research (ASOR 2013), The Australian Society for Operations Research, Adelaide, December, pp. 247–253.

Albrecht, A.R., Howlett, P.G., Pudney, P.J., Vu, X., 2011. Optimal train control: analysis of a new local optimization principle. In: Proceedings of the American Control Conference 2011.San Francisco, June 29 to July 1, 1928–1933, http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=5991057

Albrecht, A.R., Howlett, P.G., Pudney, P.J., Vu, X., 2013. Energy-efficient train control: from local convexity to global optimization and uniqueness. Automatica 49, 3072–3078. http://dx.doi.org/10.1016/j.automatica.2013.07.008

Albrecht, A.R., Howlett, P.G., Pudney, P.J., Vu, X., Zhou, P., 2014. 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.submitted for publication

Birkhoff, G., Rota, G.-C., 1989. Ordinary Differential Equations, fourth edition John Wiley & Sons.

Cheng, J., Davydova, Y., Howlett, P.G., Pudney, P.J., 1999. Optimal driving strategies for a train journey with non-zero track gradient and speed limits. IMA Journal of Mathematics Applied in Business and Industry 10, 89–115.

Cheng, J., Howlett, P.G., 1992. Application of critical velocities to the minimisation of fuel consumption in the control of trains. Automatica 28 (1), 165–169.

Cordeau, J.-F., Toth, P., Vigo, D., 1998. A survey of optimization models for train routing and scheduling. Transportation Science 32, 380–404.

Davis Jr., J.W., 1926. The tractive resistance of electric locomotives and cars. General Electric Review 29, 2–24.

Ghoseiria, K., Szidarovszkya, F., Asgharpourb, M.J., 2004. A multi-objective train scheduling model and solution. Transportation Research Part B 10, 927–952.

Girsanov, I.V., Beckman, M., Goos, G., Künzi, H.P., 1972. Lectures on Mathematical Theory of Extremum Problems. Lecture Notes in Economics and Mathematical Systems, vol. 67. Springer-Verlag.

Gong, C., Zhang, S., Zhang, F., Jiang, J., Wang, X., 2014. An integrated energy-efficient operation methodology for metro systems based on a real case of Shanghai metro line one. Energies 7, 7305–7329.

Hartl, R.F., Sethi, S.P., Vickson, R.G., 1995. A survey of the maximum principles for optimal control problems with state constraints. SIAM Review 37 (2), 181–218.

Higgins, A., Kozan, E., Ferreira, L., 1996. Optimal scheduling of trains on a single line track. Transportation Research Part B 30 (2), 147–161.

Howlett, P., 1996. Optimal strategies for the control of a train. Automatica 32 (4), 519–532.

Howlett, P., 2000. The optimal control of a train. Annals of Operations Research 98, 65–87.

Howlett, P., Jiaxing, C., 1997. Optimal driving strategies for a train on a track with continuously varying gradient. ANZIAM Journal, formerly Journal of the Australian Mathematical Society. Series B. Applied Mathematics 38, 388–410.

Howlett, P., Pudney, P., Vu, X., 2009. Local energy minimization in optimal train control. Automatica 45 (11), 2692–2698.

Howlett, P.G., Cheng, J., 1993. A note on the calculation of optimal strategies for the minimisation of fuel consumption in the control of trains. IEEE Transactions on Automatic Control 11, 1730–1734.

Howlett, P.G., Leizarowitz, A., 2001. Optimal strategies for vehicle control problems with finite control sets. Dynamics Continuous Discrete and Impulsive Systems Series B 8, 41–69.

Howlett, P.G., Milroy, I.P., Pudney, P.J., 1994. Energy-efficient train control. Control Engineering Practice 2 (2), 193–200.

Howlett, P.G., Pudney, P.J., 1995. Energy-Efficient Train Control. Advances in Industrial Control. Springer, London.

Howlett, P.G., Pudney, P.J., 1998. An optimal driving strategy for a solar powered car on an undulating road. Dynamics Continuous Discrete and Impulsive Systems 4, 553–567.

Khmelnitsky, E., 2000. On an optimal control problem of train operation. IEEE Transactions on Automatic Control 45 (7), 1257–1266.

Kraay, D., Harker, P.T., Chen, B., 1991. Pacing of trains in freight railroads: model formulation and solution. Operations Research 39 (1), 82–99.

Li, X., Lo, H.K., 2014a. An energy-efficient scheduling and speed control approach for metro rail operations. Transportation Research Part B 64, 73–89.

Li, X., Lo, H.K., 2014b. Energy minimization in dynamic train scheduling and control for metro rail operations. Transportation Research Part B 70, 269–284.

Liu, R., Golovitcher, I., 2003. Energy-efficient operation of rail vehicles. Transportation Research Part A 37, 917–932.

TTG Transportation Technology, Energymiser ®, www.ttgtransportationtechnology.com/energymiser/.

Vu, X., 2006. Analysis of Necessary Conditions for the Optimal Control of a Train. University of South Australia (PhD thesis). (See also www.amazon.co.uk/Analysis-necessary-conditions-optimal-control/dp/3639120000and search.arrow.edu.au/main/results?subject = optimality theory)

Yosida, K., 1978. Functional Analysis. Classics in Mathematics, fifth edition Springer-Verlag.

Zhou, P., Pudney, P., Howlett, P., 2013. In-train forces from energy-efficient driving strategies. In: Gaertner, P., Bowden, F., Piantadosi, J., Mobbs, K. (Eds.), 22nd National Conference of the Australian Society for Operations Research (ASOR 2013), The Australian Society for Operations Research, Adelaide, December, pp. 293–299. www.asor.org.au/conferences/asor2013/J8/zhou2.pdf

Zhou, X., Zhong, M., 2005. Train scheduling for high-speed passenger railroad planning applications. European Journal of Operational Research 167 (3), 752–771.

Zhuan, X., Xia, X., 2006. Cruise control scheduling of heavy haul trains. IEEE Transactions on Control Systems Technology 14 (4), 757–766.