Control and Cybernetics

vol. **45** (2016) No. 4

On convergence of the Monotone Structural Evolution[1]

by

Adam Korytowski[1] and Maciej Szymkat[2]

AGH University of Science and Technology, Cracow, Poland
[1] akor@agh.edu.pl, [2] msz@agh.edu.pl

Abstract: The paper studies convergence of the Monotone Structural Evolution (MSE), a computational method of optimal control. The principles of MSE are described and an expository example presents the method in action. It is then proved that under appropriate assumptions the method is convergent to the decision space stationarity conditions. Observations on finite convergence and on connections with Pontryagin's maximum principle are also provided.

Keywords: optimal control, direct computational methods, gradient optimization, monotone structural evolution

1. Introduction

The method of Monotone Structural Evolution (MSE, see Szymkat and Korytowski, 2003, 2007, Korytowski and Szymkat, 2010) is a direct method for solving numerically the optimal control problems. In the MSE algorithms, the controls are constructed by concatenation of arcs of special functions (so-called *control procedures*) taken from a predetermined finite set, the *stock*. The sequence of procedures composing a control is its *structure*. The length and contents of the control structure, the parameters of the procedures and the switching times are decision variables. The search for optimum consists of periods of gradient optimization with respect to the parameters and switching times, each period in a fixed decision space with constant dimension, separated by discrete changes of that space, called *structural changes*. The distinctive features of the MSE lie in the algorithms of the discrete part of the method, and in the construction of the stock. In every structural change, a new sequence of procedures composing the control is created, in such a way that the control does not change as a function of time. In consequence, the cost functional monotonically decreases in the course of computations. Special structural changes called *generations* are used to enrich the decision space of gradient optimization. Their purpose is to revive the search for minimum when it becomes inefficient, and

[1]Submitted: February 2016; Accepted: March 2017.

the current approximation of problem solution does not satisfy the necessary optimality conditions of Pontryagin's maximum principle. In combination with *reductions*, that is, structural changes, in which control structures (together with the decision space of gradient optimization) are simplified by eliminating their unpromising elements, the generations allow for an effective search for optimal structures.

The principal goal of this paper is to prove convergence to appropriately defined stationarity conditions for two simple versions of the MSE. The idea of the main theorem and its proof are patterned after a paper by Axelsson et al. (2008), concerning hybrid systems. This transfer of ideas was possible, because optimization in an auxiliary switched system plays a fundamental part in the MSE. Besides modifications, resulting from adapting the proof to the control theory setting and the MSE method, it was necessary to rectify some obscurities in the original reasoning, in particular in the proofs of Lemmas 4.6 and 4.7.

The paper begins with problem formulation and a presentation of the elements of the MSE. Then follow the basic MSE algorithm and an expository example of minimizing fuel consumption in car control, which shows the method in action. This example illustrates the rare property of the MSE that, unlike other computational methods, it can find the optimal control structure in a finite number of steps. Next, convergence of the basic algorithm to decision stationarity is proved under rather restrictive assumptions. A stronger theorem on infinite convergence to stationarity conditions is formulated for the MSE algorithm equipped with an Armijo-type linesearch procedure. Its proof constitutes the main result of the article. Three further examples illustrate various infinite convergence issues of the method. Observations on finite convergence and relationships between the MSE decision stationarity and the maximum principle complete the paper.

Consider the control system described by a state equation

$$\dot{x}(t) = f(x(t), u(t)), \quad t \in [0, T], \quad x(0) = x^0, \tag{1}$$

where $x(t) \in R^n$, and the control u takes values in R^m. A cost functional

$$Q(u) = q(x(T))$$

is minimized on the trajectories of (1) subject to the condition that the control is a piecewise continuous function $u : [0, T] \to U$, where U is a given nonempty set in R^m. The initial state x^0 and the horizon T are fixed. The function f is continuous in its both arguments, differentiable in the first argument, and the derivative $\partial_x f$ is continuous in its both arguments. The function q is continuously differentiable.

2. Elements of the MSE

We begin with some basic concepts of the MSE. The *control procedures* are appropriately regular functions $P : [0, T] \times R^n \times \Pi(P) \to U$, where R^n is interpreted as the state space, $\Pi(P) \subset R^{\mu(P)}$ is the set of admissible values

of the *procedure parameter*, and $\mu(P)$ denotes the dimension of the parameter. The *stock* Ξ is a predetermined, finite and nonempty set of such procedures. For an arbitrary sequence ζ, let $l(\zeta)$ denote its length. Any finite and nonempty sequence $S = (S_1, ..., S_{l(S)}) \in \Xi^{l(S)}$ is called a *control structure*. For any positive integer κ, Θ_κ denotes the set of all real sequences $\tau = (\tau_0, \tau_1, ..., \tau_\kappa)$ satisfying $0 = \tau_0 \leq \tau_1 \leq ... \leq \tau_\kappa = T$. A *decision* is defined as a triple $\xi = (S, \tau, \pi)$, where S is the *decision control structure*, $S \in \Xi^{l(S)}$, $\tau \in \Theta_{l(S)}$, and $\pi \in \Pi_S$ is the *decision parameter*, where

$$\Pi_S = \{\pi = (\pi_1, ..., \pi_{l(S)}) : \pi_i \in \Pi(S_i), \;\; i = 1, ..., l(S)\}.$$

The terms of the sequence τ are called *structural nodes*. The *length* of the decision ξ is defined by $N(\xi) = l(S)$. To each decision $\xi = (S, \tau, \pi)$, a control u_ξ and a state trajectory x_ξ *induced* by that decision are assigned, $u_\xi(t) = S_i(t, x_\xi(t), \pi_i)$, $t \in [\tau_{i-1}, \tau_i[$, $i = 1, ..., N(\xi)$, and x_ξ is the solution of (1) produced by u_ξ. Decisions are *equivalent*, if they induce the same control. The *induced cost* Σ is defined by the equality

$$\Sigma(\xi) = Q(u_\xi).$$

In the *induced optimization problem*, the minimum of Σ is sought in the set of all decisions. A decision $\hat{\xi}$ is *optimal*, if $\Sigma(\hat{\xi}) \leq \Sigma(\xi)$ for every decision ξ. If the number of optimal controls is finite, and each of them may be represented as a concatenation of a finite number of arcs of appropriately regular functions of time, state and parameters, it is always possible to choose Ξ so that the induced problem is equivalent to the original optimal control problem.

For an arbitrary control structure S, let Ω_S be the corresponding *decision space of gradient optimization*, that is, the set of all decisions of the form $\xi = (S, \tau, \pi)$ with $\tau \in \Theta_{l(S)}$ and $\pi \in \Pi_S$. Consider a situation where $\xi = (S, \tau, \pi)$ is the current approximation of the problem solution. Let $\bar{\xi} = (\bar{S}, \bar{\tau}, \bar{\pi})$ be another decision, such that $\bar{S} \neq S$ and $u_{\bar{\xi}} = u_\xi$ (whence $\Sigma(\bar{\xi}) = \Sigma(\xi)$). The *structural change* $\xi \mapsto \bar{\xi}$ denotes the transformation of ξ into $\bar{\xi}$, together with the replacement of the decision space Ω_S, which contains ξ and in which the gradient optimization ran immediately before the structural change, by the decision space $\Omega_{\bar{S}}$, to which $\bar{\xi}$ belongs. If $\bar{\xi}$ does not satisfy the stopping conditions of the algorithm, the gradient optimization is continued in the new space $\Omega_{\bar{S}}$, with $\bar{\xi}$ as the starting point.

To simplify further considerations we shall omit the explicit dependence of control procedures on time and parameters, and assume that they are functions $P : R^n \to U$. A decision thus is a pair $\xi = (S, \tau)$, where $S \in \Xi^{N(\xi)}$, $\tau \in \Theta_{N(\xi)}$, and the control induced by ξ is expressed by the formula $u_\xi(t) = S_i(x(t))$, for $t \in [\tau_{i-1}, \tau_i[$, $i = 1, ..., N(\xi)$, with x being the solution of (1) corresponding to u_ξ.

Define the functions $f_P : R^n \to R^n$, $f_P(x) = f(x, P(x))$. The state equation in the induced problem takes the form

$$\dot{x}(t) = F(t, x(t)), \;\; t \in [0, T], \;\; x(0) = x^0, \tag{2}$$

$$F(t,x) = f_{S_i}(x), \quad t \in [\tau_{i-1}, \tau_i[, \quad i = 1, ..., N(\xi).$$

We want all solutions of (2) to be well defined in $[0,T]$ and uniformly bounded. More precisely, we make the following assumption, valid throughout the paper.

Assumption 1.

(i) All solutions of the initial value problem (2) are well determined in the whole time interval $[0,T]$, and

(ii) there is a compact set $B \subset R^n$, such that all solutions of (2) lie in B and for every $P \in \Xi$, f_P is of class C^1 in some neighborhood of B.

A straightforward estimation of solution bounds for (2) yields a sufficient condition for Assumption 1 to hold.

Lemma 1. *Let $K_0 > 0$ be a constant, $\rho = \exp(K_0 T) - 1$, and $B_\rho = \{x \in R^n : ||x - x^0|| \le \rho\}$. Suppose also that for every $P \in \Xi$: (i) $||f_P(x)|| \le K_0(||x - x^0|| + 1)$ $\forall x \in B_\rho$, and (ii) f_P is of class C^1 in some neighborhood of B_ρ. Assumption 1 is then valid with $B = B_\rho$.*

The derivatives of the cost Σ with respect to structural nodes are calculated with the use of the adjoint final value problem (called *induced*)

$$\dot{\psi}(t) = -\partial_x F(t, x(t))\, \psi(t), \quad \psi(T) = -\partial q(x(T)). \tag{3}$$

Lemma 2. *There is a constant $K_1 > 0$, such that for every decision ξ the corresponding solution x of the initial value problem (2), and the corresponding solution ψ of the final value problem (3) satisfy the following relationships:*

(i) $||\, x\, ||_\infty \le K_1$, $||\psi||_\infty \le K_1$

(ii) $||x(t_2) - x(t_1)|| \le K_1 |t_2 - t_1|$, $||\psi(t_2) - \psi(t_1)|| \le K_1 |t_2 - t_1|$ $\forall\, t_1, t_2 \in [0,T]$.

Here $||\cdot||_\infty$ denotes the norm in L^∞.

Lemma 2 is an obvious consequence of Assumption 1 and the classical theorems on ordinary differential equations (see, e.g., Schättler and Ledzewicz, 2012, Appendix B).

The derivative of the cost Σ with respect to a structural node τ_i is expressed by

$$\partial_{\tau_i} \Sigma(\xi) = \psi(\tau_i)^\top (f_{S_{i+1}}(x(\tau_i)) - f_{S_i}(x(\tau_i))). \tag{4}$$

Here $\xi = (S, \tau)$, $0 < i < N(\xi)$, $\tau_{i-1} < \tau_i < \tau_{i+1}$, and x and ψ are the solutions of (1) and (3), respectively, produced by the induced control u_ξ. This well known formula comes from the classical proof of the maximum principle (see Sirisena, 1974, for an early application to gradient optimization, and Osmolovskii and Maurer, 2012, for more recent employment).

The structural changes used further are limited to *spike generations* and *reductions*. The spike generation is characterized by a function Γ which, given $P \in \Xi$, $\theta \in [0,T]$ and a decision $\xi = (S, \tau)$ with τ strictly increasing, produces

a new decision $\bar{\xi} = (\bar{S}, \bar{\tau}) = \Gamma(\xi, P, \theta)$, such that $\bar{\tau}$ is the shortest supersequence of τ with two terms equal to θ, and $\bar{S}$ is the shortest supersequence of S, such that if $\bar{\tau}_{i-1} = \bar{\tau}_i = \theta$ for some $i \in \{1, ..., N(\bar{\xi})\}$, then $\bar{S}_i = P$. The reduction ρ is a composition of two structural changes, $\rho = \rho_2 \circ \rho_1$. If $\xi = (S, \tau)$ (with an arbitrary $\tau \in \Theta_{N(\xi)}$) and $\bar{\xi} = (\bar{S}, \bar{\tau}) = \rho_1(\xi)$, then $\bar{\tau}$ is the longest strictly increasing subsequence of τ, and $\bar{S}$ is the subsequence of S, containing all those, and only those, terms S_i, $i \in \{1, ..., N(\xi)\}$, for which $\tau_{i-1} < \tau_i$. If $\bar{\xi} = (\bar{S}, \bar{\tau}) = \rho_2(\xi)$, then $\bar{S}$ is the longest subsequence of S, such that $\bar{S}_i \neq \bar{S}_{i+1}$, $i = 1, ..., N(\bar{\xi}) - 1$, and $\bar{\tau}$ is the longest subsequence of τ, such that it does not contain any term τ_i, $i \in \{1, ..., N(\xi) - 1\}$, for which $S_i = S_{i+1}$. Thus, the decisions ξ and $\rho(\xi)$ are equivalent, but the latter has no equal nodes and no equal adjacent control procedures.

A sequence of structural nodes $\tau \in \Theta_{N(\xi)}$ is called *stationary*, if

$$\tau_0 < \tau_1 < ... < \tau_{N(\xi)} \text{ and } \partial_{\tau_i} \Sigma(S, \tau) = 0, \quad i = 1, ..., N(\xi) - 1. \tag{5}$$

We say that a decision $\xi = (S, \tau)$ is *node-stationary*, if (5) holds. For the needs of this paper we take an idealistic assumption that the gradient optimization in a constant decision space Ω_S is carried on until stationarity of structural nodes is achieved.

Assumption 2. *A spike generation may be done only if conditions (5) are satisfied, that is, the decision $\xi = (S, \tau)$, which is the first argument of Γ, is node-stationary.*

Let $\bar{\xi} = (\bar{S}, \bar{\tau}) = \Gamma(\xi, P, \theta)$ and $\xi = (S, \tau)$, that is, $\bar{\xi}$ is the result of a spike generation on a node-stationary decision ξ. Define a family of decisions $\xi' = (\bar{S}, \tau')$, where τ' takes values in the intersection of a sufficiently small neighborhood of $\bar{\tau}$ with the set $\Theta_{N(\bar{\xi})}$. For a pair $P \in \Xi$, $\theta \in [0, T]$, define its *slope index* $D_{P,\theta}(\xi)$, determined by the one-sided derivatives of cost with respect to the new structural nodes, computed similarly as the derivative (4). If $\theta = 0$, then $\bar{\tau}_1 = \theta$ (by definition of $\bar{\tau}$) and

$$D_{P,0}(\xi) = \frac{\partial^+ \Sigma}{\partial \tau'_1}(\xi') \bigg|_{\xi' = \bar{\xi}} = \psi(0)^\top (f_{S_1}(x(0)) - f_P(x(0))), \tag{6}$$

and if $\tau_{i-1} < \theta \leq \tau_i$ for some $i \in \{1, ..., N(\xi)\}$, then

$$D_{P,\theta}(\xi) = -\frac{\partial^- \Sigma}{\partial \tau'_i}(\xi') \bigg|_{\xi' = \bar{\xi}} = \psi(\theta)^\top (f_{S_i}(x(\theta)) - f_P(x(\theta))) \tag{7}$$

(in that case $\bar{\tau}_i = \theta$). Here, x and ψ are the solutions of (2) and (3), respectively, induced by ξ. It easily follows from Assumptions 1 and 2 that the function $[0, T] \ni \theta \mapsto D_{P,\theta}(\xi)$ is continuous for every $P \in \Xi$. Observe that if $0 < \bar{\tau}_i = \bar{\tau}_{i+1} < T$ for a certain $i \in \{1, ..., N(\xi)\}$, then

$$-\frac{\partial^- \Sigma}{\partial \tau'_i}(\xi') \bigg|_{\xi' = \bar{\xi}} = \frac{\partial^+ \Sigma}{\partial \tau'_{i+1}}(\xi') \bigg|_{\xi' = \bar{\xi}}.$$

This is straightforward by a continuity argument, if $\theta \neq \tau_j$ for $j = 1, ..., N(\xi)$. Otherwise, Assumption 2 has to be taken into account.

Define further the *decision slope index*,

$$D(\xi) = \min_{P \in \Xi} \min_{0 \leq \theta \leq T} D_{P,\theta}(\xi).$$

It is evident that $D(\xi) \leq 0$ for every node-stationary ξ. A decision ξ is called *stationary*, if it is node-stationary and $D(\xi) = 0$. Obviously, if an optimal decision exists, then there is an optimal node-stationary decision, and every optimal node-stationary decision is stationary.

3. The basic MSE algorithm and an expository example

Conceptually, the algorithm below is a very simple implementation of the MSE ideas. Applied to the optimal control problem of Example 1, it will allow us to present the most essential elements of the MSE in action and to explain them.

Algorithm 1.

Step 0. Choose $\gamma \in]0, 1]$ and a starting decision $\bar{\xi}$. Set $k := 1$.

Step 1. Using gradient minimization and reductions, find a node-stationary decision $\xi^k = (S^k, \tau^k)$, such that $S^k_i \neq S^k_{i+1}$ for $i = 1, ..., N(\xi^k) - 1$, and

$\Sigma(\xi^k) \leq \Sigma(\bar{\xi})$, if $k = 1$,

$\Sigma(\xi^k) < \Sigma(\bar{\xi})$, if $k > 1$.

Step 2. If $D(\xi^k) = 0$, stop. Otherwise, use a spike generation to create a new decision $\bar{\xi} = \Gamma(\xi^k, P, \theta)$, such that $D_{P,\theta}(\xi^k) \leq \gamma D(\xi^k)$. Set $k := k + 1$ and return to Step 1.

Obviously, P and θ in Step 2 may be different for different values of k. For the question of choosing the parameter γ and detailed generation rules, we refer the reader to our earlier works, see also Example 1 and the examples in Section 6. It is worth mentioning here that $\gamma < 1$ allows generations with $D_{P,\theta}(\xi) > D(\xi)$, which often results in a finite convergence, while the choice of $\gamma = 1$, seemingly more effective, may lead to chattering (as in Example 3 in Section 6). It should be also remembered that the practical MSE algorithms require only approximate node-stationarity in Step 1 – more accurate near the end of optimization, and a simultaneous insertion of several spikes in Step 2 usually proves profitable.

It is evident that the performance of MSE algorithms greatly depends on the choice of the stock of control procedures Ξ. Generally, Ξ may contain any sufficiently regular functions of time, state and parameters, which prove useful in numerical approximation. However, in order to take full advantage of the possibilities offered by the MSE, it is advisable to make the stock MP-complete. For the purpose of introducing this concept, define first the *interval partition* of $[0, T]$ as any set of pairwise disjoint subintervals of $[0, T]$ with nonempty interiors, such that almost every point $t \in [0, T]$ belongs to some element of

that set. Let x_u denote the solution of (1) produced by a control u, Δ be a subinterval of $[0,T]$, and $P \in \Xi$. Denote the function $\Delta \ni t \mapsto P(x_u(t))$ by $P_\Delta(x_u)$. The set of control procedures Ξ is *MP-complete*, if and only if for every control u, satisfying the necessary optimality conditions of Pontryagin's maximum principle, there is an interval partition of $[0,T]$, possibly infinite, such that the restriction of u to any element Δ of that partition is equal to $P_\Delta(x_u)$, for some $P \in \Xi$. In particular, Ξ is MP-complete, if every such control u is induced by some decision.

It is straightforward from this definition that if the stock is MP-complete and ξ is an optimal decision, then the control induced by that decision is optimal in the original optimization problem, that is, $Q(u_\xi) \leq Q(u)$ for every piecewise continuous function $u : [0,T] \to U$.

EXAMPLE 1. Consider a car of mass m traveling along a hilly road. It is affected by the driving force v_1 produced by the engine, the force v_2 produced by friction brakes, the gravity, and the motion resistance force proportional to squared speed. The mass change, due to fuel consumption, is neglected. Denote the distance covered by the car by x_1 and its speed by x_2. It is assumed that the speed is always positive, except possibly at the initial and final time moments. Upon defining the control signals $u_1 = m^{-1}v_1$ and $u_2 = m^{-1}v_2$, we have the state equations

$$\dot{x}_1 = x_2, \;\; \dot{x}_2 = u_1 - u_2 - ax_2^2 + r(x_1), \tag{8}$$

where a is a positive constant. The function $r(x_1)$ is connected with $h(x_1)$, the height of the road above an arbitrary level, by the formula

$$r(x_1) = -\frac{g\,\partial h(x_1)}{\sqrt{1+\partial h(x_1)^2}},$$

where g is the Earth gravitational acceleration. At the initial moment of time the car is at rest, and after a given time T it should stop at a given distance y, and so

$$x_1(0) = 0, \;\; x_2(0) = 0, \tag{9}$$

$$x_1(T) = y, \;\; x_2(T) = 0. \tag{10}$$

The admissible controls are bounded

$$0 \leq u_i(t) \leq u_{im}, \;\; i = 1,2, \;\; t \in [0,T]. \tag{11}$$

The car must not violate a given speed limit V, $x_2(t) \leq V$, $t \in [0,T]$. The optimal control problem is to drive the car so that all the above conditions are satisfied and the fuel consumption is as small as possible. It is assumed that the consumed mass of fuel is proportional to the work done by the engine, and so we take the performance index in the form

$$Q(u) = \int_0^T u_1(t)x_2(t)\mathrm{d}t.$$

In order to solve this problem numerically, we use the exterior penalty approach and formulate a family of auxiliary problems, parameterized by a positive coefficient K. For a given K, the auxiliary problem is defined by the state equations (8) with the initial conditions (9), the control constraints (11) and a performance index

$$\begin{aligned} Q_K(u) &= Q(u) + \tfrac{1}{2}K\left((x_1(T)-y)^2 + x_2(T)^2 + \int_0^T (x_2(t)-V)_+^2 \mathrm{d}t\right) \\ &= x_3(T) + \tfrac{1}{2}K(x_1(T)-y)^2 + \tfrac{1}{2}Kx_2(T)^2. \end{aligned} \tag{12}$$

The additional state variable x_3 satisfies an initial value problem $\dot{x}_3 = u_1x_2 + \frac{1}{2}K(x_2 - V)_+^2$, $x_3(0) = 0$. The theory says that the solution of the auxiliary problem tends to a solution of the original optimization problem as K tends to infinity.

The preparatory stage in the MSE approach begins with the choice of the stock of control procedures Ξ. To make the stock MP-complete, we apply the maximum principle to the problem of minimizing the cost (12) on the trajectories of (8), subject to conditions (9) and (11). The Pontryagin function reads

$$H(\psi, x, u) = \psi_1x_2 + \psi_2(u_1 - u_2 - ax_2^2 + r(x_1)) - u_1x_2 - \tfrac{1}{2}K(x_2 - V)_+^2,$$

where the adjoint function ψ satisfies the adjoint set of equations

$$\begin{aligned} &\dot{\psi}_1 = -\psi_2\partial r(x_1), \;\; \psi_1(T) = K(y - x_1(T)), \\ &\dot{\psi}_2 = -\psi_1 + 2a\psi_2x_2 + u_1 + K(x_2 - V)_+, \;\; \psi_2(T) = -Kx_2(T). \end{aligned} \tag{13}$$

The switching function for the control u_1 is defined by $\phi_1(x, \psi) = \psi_2 - x_2$, and for the control u_2 by $\phi_2(x, \psi) = -\psi_2$. Any optimal control u maximizes the Pontryagin function, and so

$$u_i(t) = \left\{ \begin{array}{l} 0, \quad \text{if } \phi_i(x(t), \psi(t)) < 0 \\ u_{im}, \text{ if } \phi_i(x(t), \psi(t)) > 0 \end{array} \right\}, \;\; i = 1, 2.$$

This indicates that the stock Ξ should include the following *boundary* control procedures

$$P_1(x) = \left[\begin{array}{c} 0 \\ 0 \end{array} \right], \;\; P_2(x) = \left[\begin{array}{c} u_{1m} \\ 0 \end{array} \right], \;\; P_3(x) = \left[\begin{array}{c} 0 \\ u_{2m} \end{array} \right].$$

Obviously, control procedures with both components positive can be rejected. As it can be expected that the optimal control may have singular arcs, we shall examine the singularity conditions. The control u_1 is singular on some interval of time, if $\phi_1(x(t), \psi(t)) \equiv 0$. Assuming $u_2(t) \equiv 0$ and $x_2(t) > 0$ on that interval, and differentiating twice the identity $\phi_1[t] \equiv 0$, we obtain a state-feedback expression for the singular control, $u_{1s}(x) = ax_2^2 - r(x_1)$. The condition of singularity for u_2 on an interval of time reads $\phi_2(x(t), \psi(t)) \equiv 0$. Under the assumptions that $u_1(t) \equiv 0$ and $x_2(t) > V$, after two differentiations

of the identity $\phi_2[t] \equiv 0$ we obtain the singular control u_2 in a feedback form, $u_{2s}(x) = -ax_2^2 + r(x_1)$. The singular control arcs $u_{2s}(x)$ of the auxiliary problems tend, as $K \to \infty$, to state-constrained control arcs of the original optimal control problem, $u_2(x) = -aV^2 + r(x_1)$. In conclusion, we add two more control procedures, called *candidate singular*, to the stock Ξ

$$P_4(x) = \begin{bmatrix} \bar{u}_{1s}(x) \\ 0 \end{bmatrix}, \quad P_5(x) = \begin{bmatrix} 0 \\ \bar{u}_{2s}(x) \end{bmatrix}, \tag{14}$$

where

$$\bar{u}_{is}(x) = \left\{ \begin{array}{c} 0, \text{ if } u_{is}(x) \le 0 \\ u_{is}(x), \text{ if } 0 \le u_{is}(x) \le u_{im} \\ u_{im}, \text{ if } u_{is}(x) \ge u_{im} \end{array} \right\}, \quad i = 1, 2.$$

The stock then becomes MP-complete for the auxiliary problem.

It is worth noticing here that the adjoint equations (13) differ from the induced adjoint equations (3) on the structural time intervals, where the control procedures P_4 or P_5 are valid. Note also that substitution of (14) to the state equation may lead to discontinuity of the derivatives ∂f_P, and so to a violation of Assumption 1(ii). This can be neglected in most cases, since the continuity of ∂f_P has been assumed in order to facilitate the presentation and can be weakened. Another way to overcome this obstacle, sometimes preferred because it allows for retaining the Assumption 1(ii) and because of a better rate of convergence, are special *saturation generations*, performed after each line-search ending with saturation of a candidate singular procedure. They consist in adding a new structural node at every time moment, different from the already existing nodes, at which the function $t \mapsto \partial_x F(t, x(t))$ is discontinuous, together with introducing the appropriate new elements in the control structure. In accordance with the general rules of structural changes, a saturation generation does not change the control as a function of time, and the new nodes fulfill the stationarity condition (5). A saturation generation can be followed by a few steps of non-gradient optimization, involving only those structural nodes, at which the adjacent control procedures take equal values.

For calculations we take $g = 9.81$ m/s^2, $y = 4000$ m, $T = 240$ s, $a = 0.001$, $u_{1m} = 2$ m/s^2, $u_{2m} = 4$ m/s^2, $V = 30$ m/s. The function r is a C^1 piecewise polynomial, constructed by means of the MATLAB function `pchip`

$$r(x_1) = r_{1i}(x_1 - y_i)^3 + r_{2i}(x_1 - y_i)^2 + r_{3i}, \quad y_i \le x_1 \le y_{i+1}, \quad i = 1, ..., 7,$$

where $y_8 = y$ and the values of y_i, r_{1i}, r_{2i}, r_{3i}, $i = 1, ..., 7$, are given in the table below:

i	1	2	3	4	5	6	7
y_i	0	100	400	700	1000	2000	2300
$10^5 r_{1i}$	0	1/135	0	−1/135	0	−1/108	0
$10^4 r_{2i}$	0	−1/3	0	1/3	0	5/12	0
r_{3i}	0	0	−1	−1	0	0	5/4

The inserted control procedure P and the insertion point θ in spike generations are chosen in a rather complex way[1]. First, for $i = 1, ..., N(\xi)$ find $p_i \in \Xi$ and $t_i \in [\tau_{i-1}, \tau_i]$, satisfying

$$D_{p_i,t_i}(\xi) = \min\{D_{p,\,t}(\xi) : p \in \Xi,\ t \in [\tau_{i-1}, \tau_i]\}.$$

Let $I_\gamma = \{i \in \{1, ..., N(\xi)\} : D_{p_i,t_i}(\xi) \le \gamma D(\xi)\}$. Then, for every $i \in I_\gamma$ find $\bar{p}_i \in \Xi$, such that

$$D_{\bar{p}_i,t_i}(\xi) = \max\{D_{p,t_i}(\xi) : p \in \Xi,\, D_{p,t_i}(\xi) \le \gamma D(\xi)\}.$$

Finally, $P = \bar{p}_j$ and $\theta = t_j$, where $j \in I_\gamma$ and $D_{\bar{p}_j,t_j}(\xi) = \min\{D_{\bar{p}_i,t_i}(\xi) :\ i \in I_\gamma\}$.

We begin computations with the penalty coefficient $K = 10$. In Step 0, set $\gamma := 0.02$, $\bar{\xi} := (\bar{S}, \bar{\tau})$, $\bar{S} := (P_2, P_1)$, $\bar{\tau} := (0,\ 45,\ T)$. The results of further calculations for $K = 10$ are summarized in the table below, where $\Sigma(\xi) = Q_K(u_\xi)$, and all other symbols are as in the description of Algorithm 1 and in the rule of choice of P and θ.

k	$\Sigma(\xi^k)$	$D(\xi^k)$	S^k	j	P	θ	$D_{P,\theta}(\xi^k)$
1	9880.09	−2052.29	(P_2, P_1)	2	P_5	222.354	−52.94
2	2954.59	−450.15	(P_2, P_1, P_5)	3	P_3	240	−450.15
3	1405.37	−44.57	(P_2, P_1, P_5, P_3)	1	P_4	16.8505	−11.73
4	1018.30	−0.00002	$(P_2, P_4, P_1, P_5, P_3)$				

The corresponding node vectors are as follows:

$$\tau^1 = (0,\ 31.35736,\ 240),$$
$$\tau^2 = (0,\ 37.85850,\ 121.93994,\ 240),$$
$$\tau^3 = (0,\ 31.97458,\ 194.21181,\ 230.14511,\ 240),$$
$$\tau^4 = (0,\ 7.06500,\ 132.06677,\ 193.47539,\ 230.15353,\ 240).$$

The value of the slope index $D(\xi^4)$ is zero within the limits of computational accuracy (its negative value does not mean that the algorithm can be continued with a new spike generation) and so $\xi^4 = (S^4, \tau^4)$ is recognized as an optimal decision for $K = 10$. The continuation of the penalty method with greater values of K does not change the optimal control structure, and the changes of the plots in Figs. 1d, 2 and 3 are negligible. The controls u_1 and u_2, induced by ξ^k, $k = 1, 2, 3$, are shown in the upper parts of Figs. 1a, 1b and 1c, respectively; the candidate singular controls $\bar{u}_{1s}$ and $\bar{u}_{2s}$ are plotted with thin lines. The negative arcs of the functions $D_i(t) = D_{P_i,t}(\xi^k)$, $i = 1, ..., 5$, are shown in the lower parts of the figures. The vertical bold dotted lines ended with dots indicate the spikes inserted in each iteration. They are prolonged downwards with thin solid lines to help understand their connection with the slope indices.

[1]It is not difficult to propose simpler rules for choosing P and θ, which would also ensure finite convergence, however, the number of iterations could then be greater.

The optimal controls as functions of time are shown in Fig. 1d together with the normalized switching functions $\bar{\phi}_i(t) = \phi_i[t]/\max_z |\,\phi_i[z]\,|$, $i = 1, 2$. It can be seen that the necessary optimality conditions of the maximum principle are satisfied within computational accuracy. Figure 2 presents plots of the optimal controls as functions of the distance x_1, and the road profile described by the function $h(x_1)$. The phase plots of the optimal state trajectories $x_2(x_1)$ and $x_3(x_1)$ are depicted in Fig. 3.

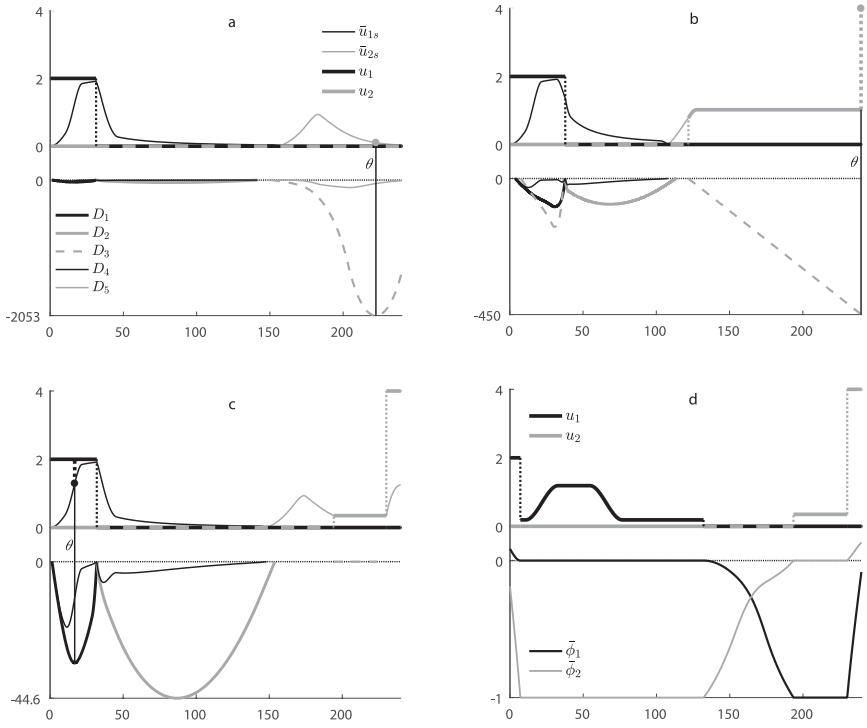

Figure 1. Results of optimization and spike generations in iterations 1 (a), 2 (b), and 3 (c), and the optimal control with switching functions (d) (time on horizontal axes)

All computations were performed with the use of a fully implicit 5th order, 3 stage Radau IIa/Radau Ia adjoint pair of ODE solvers, with the relative step accuracy of 10^{-14}.

4. On infinite convergence of the MSE algorithms

Despite the simplicity of Algorithm 1, we can only give a rather weak characterization of its convergence. In order to prove a stronger theorem it will

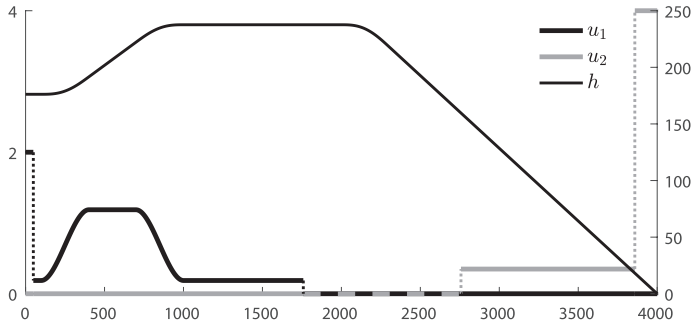

Figure 2. Optimal controls u_1 and u_2, and the road profile h as functions of x_1; control values on the left, values of h on the right

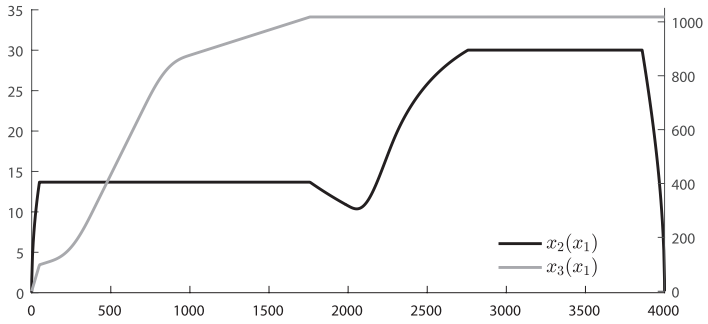

Figure 3. Optimal phase trajectories $x_2(x_1)$ (left scale) and $x_3(x_1)$ (right scale)

be necessary to equip the MSE algorithm with an additional Armijo linesearch (similar to the procedure applied in Axelsson et al., 2008).

Lemma 3. *Assume that*

(i) Algorithm 1 is infinite,

(ii) there is an increasing function $\zeta : R_- \to R_-$*, such that* $\Sigma(\xi^{k+1}) - \Sigma(\xi^k) \leq \zeta(D(\xi^k))$ *for every* k*.*

Then, $D(\xi^k) \to 0$ *as* $k \to \infty$*.*

Proof. Assume that the lemma is false. Then, there exist a real number $\delta < 0$ and an infinite strictly increasing sequence of positive integers k_i, $i = 1, 2, ...$, such that $D(\xi^{k_i}) < \delta$, $i = 1, 2, ...$ By assumption, $\Sigma(\xi^{k_i+1}) - \Sigma(\xi^{k_i}) \leq \zeta(\delta)$, where $\zeta(\delta)$ is a negative real number, independent of the iteration number k. Since the cost functional Σ is lower bounded (by virtue of Assumption 1), we have come to a contradiction. ■

To fulfill the assumption (ii) of Lemma 3, we shall make an additional assumption regarding Step 1, or, alternatively, modify Algorithm 1 by adding a new step. Before stating the theorems on convergence, we need some definitions and lemmas. We say that Step 1 of Algorithm 1 is *complete* for a given k, if the point τ^k determined in that step is a minimizer of the function $\tau \mapsto \Sigma(S^k, \tau)$ in $\Theta_{l(S^k)}$.

The following construction plays an important part in the sequel. Consider a spike generation $\xi \mapsto \bar{\xi}$, which transforms a decision $\xi = (S, \tau)$ into $\bar{\xi} = (\bar{S}, \bar{\tau}) = \Gamma(\xi, P, \theta)$ by inserting a control procedure P at a time moment θ. In accordance with Assumption 2, the sequence τ is stationary. Assume also that $D_{P,\,\theta}(\xi) < 0$. Given ξ and $\bar{\xi}$, define a broken line $[0, \infty[\ni \lambda \mapsto c(\lambda) = (c_0(\lambda), c_1(\lambda), ..., c_{l(c(\lambda))}(\lambda))$, whose successive segments lie in the real vector spaces with decreasing dimensions. Assume $c(0) = \bar{\tau}$. For $\lambda > 0$, $c(\lambda)$ is the shortest increasing sequence comprising all elements of the set

$$\{\max(0, \theta - \lambda), \min(\theta + \lambda, T)\} \cup \{\tau_i \in [0, \theta - \lambda] \cup [\theta + \lambda, T] : \ i \in \{0, ..., N(\xi)\}\}.$$

Let $z(\lambda)$, for $\lambda \geq 0$ denote the subsequence of $\bar{S}$ which is a concatenation of the sequences

$$(S_i : \tau_{i-1} < \theta - \lambda, \ i \in \{1, ..., N(\xi)\}), (P), \text{ and } (S_i : \tau_i > \theta + \lambda, \ i \in \{1, ..., N(\xi)\}).$$

Note that the first segment of the broken line c has the steepest descent direction. Let further $\xi_\lambda = (z(\lambda), c(\lambda))$ and $h(\lambda) = \Sigma(\xi_\lambda)$. The following lemma gives an estimate of the induced cost Σ on the broken line c. As the estimation is lengthy and technical, the proof has been moved to a separate Section 5.

LEMMA 4. *There is a positive constant $L > 0$, independent of ξ, P and θ, such that*

$$h(\lambda) \leq h(0) + D_{P,\,\theta}(\xi)\lambda + \tfrac{1}{2}L\lambda^2$$

for every $\lambda \in [0, \lambda_{\max}]$, where $\lambda_{\max} = \min(-2L^{-1}D_{P,\,\theta}(\xi), \ \frac{1}{2}T)$.

Recall that decisions are equivalent, if they induce identical controls. The next lemma directly follows from that definition.

LEMMA 5. *To every decision $\xi = (S, \tau)$ there is an equivalent decision $\bar{\xi} = (\bar{S}, \bar{\tau})$, such that $\bar{S}$ and $\bar{\tau}$ are subsequences of S and τ, respectively, $\bar{\tau}$ is strictly increasing, and $\bar{S}_i \neq \bar{S}_{i+1}$ for $i = 1, ..., N(\bar{\xi}) - 1$.*

We can now prove that completeness of Step 1 is sufficient for the assumption (ii) of Lemma 3 to hold.

THEOREM 1. *Assume that Step 1 of Algorithm 1 is complete for every k, and Algorithm 1 is infinite. Then $D(\xi^k) \to 0$ as $k \to \infty$.*

PROOF. It is straightforward from Lemmas 4 and 5 that the minimum of the function $\tau \mapsto \Sigma(S^k, \tau)$ in $\Theta_{l(S^k)}$ is not greater than $\Sigma(\xi^k) + \zeta_0(D_{P,\theta}(\xi^k))$, where

$$\zeta_0(v) = \begin{cases} -\frac{1}{2}L^{-1}v^2, & -L^{-1}v \le \frac{1}{2}T \\ \frac{1}{2}vT + \frac{1}{4}LT^2, & -L^{-1}v \ge \frac{1}{2}T \end{cases}$$

and L is a positive constant, independent of k. As $\zeta_0 : R_- \to R_-$ is an increasing function and $D_{P,\theta}(\xi^k) \le \gamma D(\xi^k)$, then $\Sigma(\xi^{k+1}) \le \Sigma(\xi^k) + \zeta_0(\gamma D(\xi^k))$. The function $\zeta(v) = \zeta_0(\gamma v)$ is also increasing and the assumptions of Lemma 3 are fulfilled. ■

The completeness of Step 1, or even a good approximation of it, seldom can be guaranteed in practical problems. So far, the broken line c has only been used to prove the existence of ζ. We shall now show that equipping the algorithm with an Armijo-type linesearch along c allows us to skip the assumption of completeness.

The *Armijo linesearch* is an algorithm for finding better cost values $h(\lambda)$ on the broken line c, based on a modification of the well known Armijo condition. Assume a constant $\alpha \in]0, 1[$ and a strictly decreasing, infinite sequence of real numbers $(\lambda_j,\ j = 1, 2, ...) \subset]\,0, \frac{1}{2}T]$, converging to zero. The result of the search for a minimum on c is defined as a decision $\xi_\Lambda = (z(\Lambda), c(\Lambda))$, where

$$\Lambda = \max\{\lambda_j,\ j = 1, 2, ...\ :\ h(\lambda_j) - h(0) \le \alpha\lambda_j D_{P,\theta}(\xi)\}. \tag{15}$$

A simple MSE algorithm with the Armijo linesearch is given below.

ALGORITHM 2.

Step 0. Choose $\gamma \in]0, 1]$ and a starting decision $\hat{\xi}$. Set $k := 1$.

Step 1. Using gradient minimization and reductions, find a node-stationary decision $\xi^k = (S^k, \tau^k)$, such that $\Sigma(\xi^k) \le \Sigma(\hat{\xi})$ and $S_i^k \ne S_{i+1}^k$ for $i = 1, ..., N(\xi^k)-1$.

Step 2. If $D(\xi^k) = 0$, stop. Otherwise, use a spike generation to create a new decision $\bar{\xi} = \Gamma(\xi^k, P, \theta)$, such that $D_{P,\theta}(\xi^k) \le \gamma D(\xi^k)$.

Step 3. Determine the broken line c as earlier in this section, putting $\xi = \xi^k$. Using the Armijo linesearch find ξ_Λ and set $\hat{\xi} := \xi_\Lambda$. Next, set $k := k + 1$ and return to Step 1.

The following theorem is the main result of this paper.

THEOREM 2. *Let ξ^k, $k = 1, 2, ...$, be the sequence of decisions produced in Step 1 of successive iterations of Algorithm 2. If that sequence is infinite, then* $\lim_{k\to\infty} D(\xi^k) = 0$.

PROOF. Assume that the sequence ξ^k, $k = 1, 2, ...$, is infinite, and apply Lemma 4 to Step 3 of iteration k of Algorithm 2. The cost estimate in Lemma 4 $a(\lambda) = \Sigma(\xi^k) + D_{P,\theta}(\xi^k)\lambda + \frac{1}{2}L\lambda^2$, and the straight line $b(\lambda) = \Sigma(\xi^k) + \alpha D_{P,\theta}(\xi^k)\lambda$ intersect at $\lambda = 2(\alpha - 1)L^{-1}D_{P,\theta}(\xi^k)$. Define $\hat{\lambda}(v) = \max\{\lambda_j,\ j = 1, 2, ...\ :\ \lambda_j \le 2(\alpha - 1)L^{-1}v\}$ for $v < 0$. At the point Λ produced

by the Armijo linesearch we have $\Sigma(\xi_\Lambda) \leq b(\Lambda) \leq b(\hat{\lambda}(D_{P,\theta}(\xi^k)))$ (because $\hat{\lambda}(D_{P,\theta}(\xi^k)) \leq \Lambda$). Thus, $\Sigma(\xi^{k+1}) \leq \Sigma(\xi^k) + \alpha D_{P,\theta}(\xi^k)\hat{\lambda}(D_{P,\theta}(\xi^k))$. As $\zeta_0(v) = \alpha\, v\hat{\lambda}(v)$ is an increasing function $R_- \to R_-$ independent of the iteration number k, so is $\zeta(v) = \zeta_0(\gamma v)$, and the assumptions of Lemma 3 are satisfied. ■

5. Proof of Lemma 4

Our aim is to find an upper bound for $h(\lambda) = \Sigma(\xi_\lambda)$. To that end, we first estimate the right-hand cost derivative on c. Denote by x_λ the solution of (1) induced by ξ_λ, and by ψ_λ the corresponding solution of (3). Notice that x_0 and ψ_0 are, respectively, the solutions of (1) and (3), produced by the control u_ξ. The right derivative of h at zero is given by

$$\partial^+ h(0) = \begin{cases} 2D_{P,\theta}(\xi), & \text{if } 0 < \theta < T \\ D_{P,\theta}(\xi), & \text{if } \theta = 0 \text{ or } \theta = T \end{cases}$$

(compare (6) and (7)). We shall need the right derivative $\partial^+ h(\lambda)$ for an arbitrary $\lambda \geq 0$, which can be computed analogously to (4). It will be convenient to write the derivative as a sum, $\partial^+ h(\lambda) = d_-(\lambda) + d_+(\lambda)$, where

$$d_-(\lambda) = \begin{cases} \psi_\lambda(\theta-\lambda)^\top (f_{S_{r(\lambda)}}(x_\lambda(\theta-\lambda)) - f_P(x_\lambda(\theta-\lambda))), & \theta - \lambda > 0 \\ 0, & \theta - \lambda \leq 0, \end{cases}$$

$$d_+(\lambda) = \begin{cases} \psi_\lambda(\theta+\lambda)^\top (f_{S_{s(\lambda)}}(x_\lambda(\theta+\lambda)) - f_P(x_\lambda(\theta+\lambda))), & \theta + \lambda < T \\ 0, & \theta + \lambda \geq T, \end{cases}$$

$$r(\lambda) = \min\{\, i \in \{1, ..., N(\xi)\} : \tau_i \geq \theta - \lambda\} \text{ for } \theta - \lambda > 0,$$

$$s(\lambda) = \min\{\, i \in \{1, ..., N(\xi)\} : \tau_i > \theta + \lambda\} \text{ for } \theta + \lambda < T.$$

We shall show that $\partial^+ h(\lambda)$ is bounded by a linear function of λ. The next lemma gives an estimate for the derivative of cost on the broken line c, whose construction follows a spike generation $\xi \to \bar{\xi}$ with $\xi = (S, \tau)$ and $\bar{\xi} = (\bar{S}, \bar{\tau}) = \Gamma(\xi, P, \theta)$. By Assumption 2, the sequence τ satisfies (5).

LEMMA 6. *There is a constant $L > 0$, independent of ξ, P and θ, such that*
(i) $|\, d_-(\lambda) - d_-(0)\,| \leq L\lambda$ *for every* $\lambda \in [0, \theta[$,
(ii) $|\, d_+(\lambda) - d_+(0)\,| \leq L\lambda$ *for every* $\lambda \in [0, T - \theta[$.

PROOF. We only prove (i), because for (ii) the argument is similar. Suppose that $0 \leq \lambda < \theta$. Then, $d_-(\lambda) - d_-(0) = V_1 - V_2$, where

$$V_1 = \psi_\lambda(\theta-\lambda)^\top f_{S_{r(\lambda)}}(x_\lambda(\theta-\lambda)) - \psi_0(\theta)^\top f_{S_{r(0)}}(x_0(\theta)),$$

$$V_2 = \psi_\lambda(\theta-\lambda)^\top f_P(x_\lambda(\theta-\lambda)) - \psi_0(\theta)^\top f_P(x_0(\theta)).$$

We first find an estimate for V_2. Put $a = x_\lambda(\theta-\lambda) - x_0(\theta)$ and $b = \psi_\lambda(\theta-\lambda) - \psi_0(\theta)$. From Assumption 1 and the classical theorems on the dependence of

solutions to ordinary differential equations on parameters, it follows that there is a constant $K_2 > 0$, such that $||x_\lambda - x_0||_\infty \leq K_2\lambda$ and $||\psi_\lambda - \psi_0||_\infty \leq K_2\lambda$ for every $\lambda \geq 0$. Therefore, by Lemma 2

$$||a|| \leq ||x_\lambda(\theta - \lambda) - x_\lambda(\theta)|| + ||x_\lambda(\theta) - x_0(\theta)|| \leq (K_1 + K_2)\lambda,$$

and in a similar manner, $||b|| \leq (K_1 + K_2)\lambda$. Further,

$$f_P(x_\lambda(\theta - \lambda)) = f_P(x_0(\theta) + a) = f_P(x_0(\theta)) + \partial f_P(x_0(\theta))^\top a + o(a),$$

o being a common symbol for all error terms of order higher than one. After a short calculation,

$$V_2 = b^\top f_P(x_\lambda(\theta - \lambda)) + \psi_0(\theta)^\top \partial f_P(x_0(\theta))^\top a + o(a).$$

The right-hand side is of order at least one with respect to a and b, and so there exists a constant $L_2 > 0$, such that $|V_2| \leq L_2\lambda$. The estimation of V_1 is more complicated. Write it in the form

$$V_1 = V_{1,1} + V_{1,2} + V_{1,3},$$

$$V_{1,1} = \psi_\lambda(\theta - \lambda)^\top f_{S_{r(\lambda)}}(x_\lambda(\theta - \lambda)) - \psi_0(\theta)^\top f_{S_{r(\lambda)}}(x_0(\theta)),$$

$$V_{1,2} = \psi_0(\theta)^\top f_{S_{r(\lambda)}}(x_0(\theta)) - \psi_0(\bar\tau_{r(\lambda)})^\top f_{S_{r(\lambda)}}(x_0(\bar\tau_{r(\lambda)})),$$

$$V_{1,3} = \psi_0(\bar\tau_{r(\lambda)})^\top f_{S_{r(\lambda)}}(x_0(\bar\tau_{r(\lambda)})) - \psi_0(\theta)^\top f_{S_{r(0)}}(x_0(\theta)).$$

Repeating the argument used to estimate V_2 it can be shown that there is a constant $L_{1,1} > 0$, such that $|V_{1,1}| \leq L_{1,1}\lambda$. In a similar way it is proved that $|V_{1,2}| \leq L_{1,2}|\theta - \bar\tau_{r(\lambda)}|$, for some constant $L_{1,2} > 0$. As it easily follows from the definition that $0 \leq \theta - \bar\tau_{r(\lambda)} \leq \lambda$, we obtain $|V_{1,2}| \leq L_{1,2}\lambda$. Let us now pass to the estimation of $V_{1,3}$. For $r(\lambda) = r(0)$, $V_{1,3} = 0$. For $r(\lambda) < r(0)$,

$$\begin{aligned} V_{1,3} &= \sum_{i=r(\lambda)}^{r(0)-1} \left(\psi_0(\bar\tau_i)^\top f_{S_i}(x_0(\bar\tau_i)) - \psi_0(\bar\tau_{i+1})^\top f_{S_{i+1}}(x_0(\bar\tau_{i+1})) \right) \\ &= \sum_{i=r(\lambda)}^{r(0)-1} (W_{1,3,i} + Z_{1,3,i}), \end{aligned}$$

where

$$W_{1,3,i} = \psi_0(\bar\tau_i)^\top \left(f_{S_i}(x_0(\bar\tau_i)) - f_{S_{i+1}}(x_0(\bar\tau_i)) \right),$$

$$Z_{1,3,i} = \psi_0(\bar\tau_i)^\top f_{S_{i+1}}(x_0(\bar\tau_i)) - \psi_0(\bar\tau_{i+1})^\top f_{S_{i+1}}(x_0(\bar\tau_{i+1})).$$

For $i = r(\lambda), ..., r(0)-1$, we have $\bar\tau_i = \tau_i$. Hence, by virtue of the assumption that τ is stationary, $W_{1,3,i} = \psi_0(\tau_i)^\top \left(f_{S_i}(x_0(\tau_i)) - f_{S_{i+1}}(x_0(\tau_i)) \right) = \partial_{\tau_i}\Sigma(\xi) = 0$. Consider now the term $Z_{1,3,i}$. Reasoning analogously as in the case of $V_{1,2}$,

it can be shown that there exists a constant $L_{1,3} > 0$, such that $|Z_{1,3,i}| \leq L_{1,3}(\bar{\tau}_{i+1} - \bar{\tau}_i)$ for $i = r(\lambda), ..., r(0) - 1$. Therefore

$$\left| \sum_{i=r(\lambda)}^{r(0)-1} Z_{1,3,i} \right| \leq \sum_{i=r(\lambda)}^{r(0)-1} |Z_{1,3,i}| \leq L_{1,3} \sum_{i=r(\lambda)}^{r(0)-1} (\bar{\tau}_{i+1} - \bar{\tau}_i) = L_{1,3}(\bar{\tau}_{r(0)} - \bar{\tau}_{r(\lambda)}) \leq L_{1,3}\lambda.$$

Consequently, $|V_{1,3}| \leq L_{1,3}\lambda$ and further, for $L_1 = L_{1,1}+L_{1,2}+L_{1,3}$, $|V_1| \leq L_1\lambda$. Putting $L_- = L_1 + L_2$ yields that $|d_-(\lambda) - d_-(0)| \leq L_-\lambda$ for every $\lambda \in [0, \theta[$. A similar argument leads to the conclusion that $|d_+(\lambda) - d_+(0)| \leq L_+\lambda$ for some $L_+ > 0$ and every $\lambda \in [0, T - \theta[$. The claim of Lemma 6 is obtained with $L = \max(L_-, L_+)$. ■

Now we can estimate the cost on c. By Lemma 6, $d_-(\lambda) \leq D_{P,\theta}(\xi) + L\lambda$ for $\lambda \in [0, \theta[$ and $d_-(\lambda) = 0$ for $\lambda \geq \theta$. Similarly, $d_+(\lambda) \leq D_{P,\theta}(\xi) + L\lambda$ for $\lambda \in [0, T - \theta[$ and $d_+(\lambda) = 0$ for $\lambda \geq T - \theta$. As

$$h(\lambda) = h(0) + \int_0^\lambda \partial^+ h(s)\, \mathrm{d}s,$$

we obtain

$$h(\lambda) \leq h(0) + 2D_{P,\theta}(\xi)\lambda + L\lambda^2, \quad 0 \leq \lambda \leq \lambda_1,$$

$$h(\lambda) \leq h(0) + D_{P,\theta}(\xi)(\lambda_1 + \lambda) + \tfrac{1}{2}L(\lambda_1^2 + \lambda^2), \quad \lambda_1 \leq \lambda \leq \lambda_2,$$

with $\lambda_1 = \min(\theta, T - \theta)$ and $\lambda_2 = \max(\theta, T - \theta)$. Since $\lambda_2 \geq \frac{1}{2}T$, the claim of Lemma 4 follows.

6. Further examples

Example 2 (non MP-complete stock). Consider the fuel consumption problem of Example 1. Once more we apply Algorithm 1 to the auxiliary problem with the penalty coefficient $K = 10$, but this time with a non MP-complete stock $\Xi = \{P_1, P_2, P_3, P_4\}$. The generation rules, the starting decision and the parameter γ are as before. With such a stock, it is still possible to construct optimizing sequences of decisions (since all the boundary procedures are present), but the finite convergence cannot be achieved. Our aim is to estimate the rate of convergence. Starting from iteration 6, all decisions $\xi^k = (S^k, \tau^k)$ exhibit a characteristic pattern. The induced controls u^k, $k = 6, 7, ...$, do not differ much outside the time interval $[\tau_3^k, \tau_{2k-3}^k[$, that is, between the fourth and the penultimate structural nodes. In that interval u_1^k is identically zero, and u_2^k is of bang-bang type with $2k - 6$ switches (see Fig. 4, where the controls u^6 (a) and u^{18} (b) are depicted). More precisely, for every $k \geq 6$ we have $l(S^k) = 2(k - 1)$, and $S_1^k = P_2$, $S_2^k = P_4$, $S_{2i-1}^k = P_1$, $S_{2i}^k = P_3$, $i = 2, ..., k - 1$.

The control u_2^k and the corresponding trajectory x_2 are periodic in $[\tau_3^k, \tau_{2k-3}^k[$ with a period $\delta_k = \tau_5^k - \tau_3^k$. To prove it, assume additionally that $x_2(t) \geq V$ for

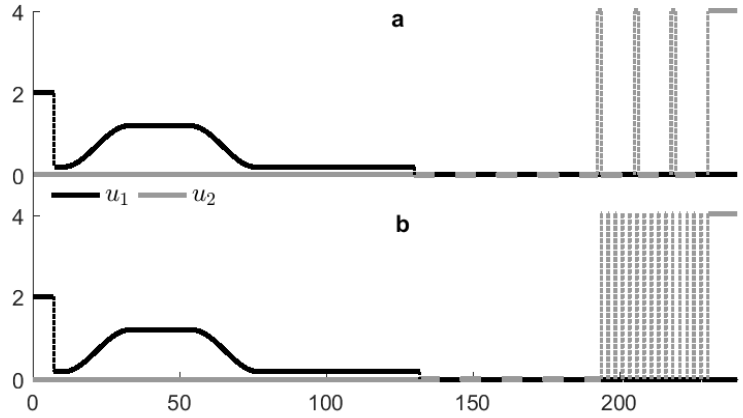

Figure 4. The controls u^6(a) and u^{18}(b) (time on horizontal axes)

every $t \in [\tau_3^k, \tau_{2k-3}^k[$, which is true for all sufficiently large k's (the proof for an arbitrary k is similar, but the details are more complicated). Let ψ denote the corresponding solution of the induced adjoint system (3). By the stationarity of τ^k (see (5)), we have $\psi_2(\tau_i^k) = 0$, $i = 3, ..., 2k-3$. Note that in the time intervals where $r(x_1(t)) = \text{const}$, also $\psi_1 = \text{const}$. Then, by straightforward substitution it can be verified that for every $i = 2, ..., k-2$, and $t \in [\tau_{2i-1}^k, \tau_{2i+1}^k]$

$$\psi_2(t) = K(x_2(t) - x_2(\tau_{2i}^k)) \frac{\frac{1}{2}(x_2(t) + x_2(\tau_{2i}^k)) - W_k}{r_{37} - u_2(t) - a x_2(t)^2},$$

where $W_k = V + K^{-1}\psi_1$, $u_2(t) = u_{2m}$ for $t \in [\tau_{2i-1}^k, \tau_{2i}^k[$, $u_2(t) = 0$ for $t \in [\tau_{2i}^k, \tau_{2i+1}^k[$. Hence $x_2(\tau_{2i-1}^k) = x_2(\tau_{2i+1}^k)$, $i = 2, ..., k-2$, and $x_2(\tau_{2i}^k) = x_2(\tau_{2i+2}^k)$, $i = 2, ..., k-3$. This, and the alternate monotonicity of x_2, imply the equalities $\tau_4^k - \tau_3^k = ... = \tau_{2k-4}^k - \tau_{2k-5}^k$ and $\tau_5^k - \tau_4^k = ... = \tau_{2k-3}^k - \tau_{2k-4}^k$, and the periodicity follows.

The slope indices $D_{P,\theta}(\xi^k)$ for $k \geq 6$ are negative for all pairs (P, θ) of the form (P_1, θ), $\theta \in]\tau_{2i-1}, \tau_{2i}[$, and (P_3, θ), $\theta \in]\tau_{2i}, \tau_{2i+1}[$, where $i = 2, ..., k-2$. For all other pairs, $D_{P,\theta}(\xi^k) \geq 0$. The indices $D_{P_1,\theta}(\xi^k)$ and $D_{P_3,\theta}(\xi^k)$ as functions of $\theta \in [\tau_3^k, \tau_{2k-3}^k[$ are periodic with the period δ_k. Thus, the decision slope index $D(\xi^k)$ is equal to the lesser of the two numbers: $\min\{D_{P_1,\theta}(\xi^k) : \theta \in [\tau_3^k, \tau_4^k]\}$ and $\min\{D_{P_3,\theta}(\xi^k) : \theta \in [\tau_4^k, \tau_5^k]\}$. For the assumed values of parameters the latter is always smaller, and so

$$D(\xi^k) = \min\{D_{P_3,\theta}(\xi^k) : \theta \in [\tau_4^k, \tau_5^k]\} = u_{2m} \min\{\psi_2(\theta) : \theta \in [\tau_4^k, \tau_5^k]\}.$$

To estimate the rate of convergence of the MSE, assume that k is sufficiently large and write

$$\psi_2(\theta) = \dot{\psi}_2(\tau_4^k)(\theta - \tau_4^k) + \tfrac{1}{2}\ddot{\psi}_2(\tau_4^k \pm)(\theta - \tau_4^k)^2 + o(\delta_k^2), \quad \theta \in [\tau_3^k, \tau_5^k],$$

where 'τ_4^k-' is valid for $\theta \in [\tau_3^k, \tau_4^k]$ and 'τ_4^k+' for $\theta \in [\tau_4^k, \tau_5^k]$. Here

$$\dot{\psi}_2(\tau_4^k) = K(x_2(\tau_4^k) - W_k),$$
$$\ddot{\psi}_2(\tau_k^4-) = K(r_{37} - u_{2m} - aW_k^2 + a(x_2(\tau_4^k) - W_k)^2),$$
$$\ddot{\psi}_2(\tau_k^4+) = K(r_{37} - aW_k^2 + a(x_2(\tau_4^k) - W_k)^2).$$

Using the stationarity condition (5) we easily obtain

$$\tau_4^k - \tau_3^k = 2\frac{x_2(\tau_4^k) - W_k}{r_{37} - aW_k^2 - u_{2m}} + o(\delta_k), \quad \tau_5^k - \tau_4^k = 2\frac{x_2(\tau_4^k) - W_k}{aW_k^2 - r_{37}} + o(\delta_k).$$

Hence

$$x_2(\tau_4^k) - W_k = \frac{(u_{2m} + aW_k^2 - r_{37})(aW_k^2 - r_{37})}{2\,u_{2m}}\delta_k + o(\delta_k)$$

and

$$\begin{aligned} D(\xi^k) &= -u_{2m}\frac{\dot{\psi}_2(\tau_4^k)^2}{2\ddot{\psi}_2(\tau_4^k+)} + o(\delta_k^2) \\ &= \frac{K(u_{2m} + aW_k^2 - r_{37})^2(aW_k^2 - r_{37})}{8\,u_{2m}}\delta_k^2 + o(\delta_k^2). \end{aligned}$$

Since

$$\delta_k = \frac{\tau_{2k-3}^k - \tau_3^k}{2(k-3)},$$

and W_k and $\tau_{2k-3}^k - \tau_3^k$ have finite limits W and Δ, respectively, as $k \to \infty$, then

$$D(\xi^k) = \frac{K(u_{2m} + aW^2 - r_{37})^2(aW^2 - r_{37})\,\Delta^2}{32\,u_{2m}(k-3)^2} + o(k^{-2}), \quad k > 3.$$

The first term in the right-hand side is a good approximation of $D(\xi^k)$ if k is so large that $x_2(t) \geq V$ for every $t \in [\tau_3^k, \tau_{2k-3}^k]$.

Example 3 (the role of parameter γ). Consider a scalar system, described by a state equation $\dot{x}(t) = u(t)$, $t \in [0, T]$, with $T = 2$ and $x(0) = 1$. A cost functional

$$Q(u) = \tfrac{1}{2}\int_0^T x(t)^2 \mathrm{d}t$$

is to be minimized subject to a control constraint $|u(t)| \leq 1$. This problem can be transformed (similarly as in Section 3) to the form given in Section 1. Pontryagin's maximum principle then yields the necessary optimality conditions for a control u and the corresponding trajectory x :

$$u(t) = \begin{cases} -1, & \text{if } \psi(t) < 0 \\ +1, & \text{if } \psi(t) > 0\,, \end{cases}$$

where the adjoint function ψ is a solution of the final value problem $\dot{\psi}(t) = x(t)$, $\psi(T) = 0$. In every open interval where $\psi(t) \equiv 0$, the optimal control is identically zero. Indeed, we then have $\psi(t) \equiv 0 \equiv \dot{\psi}(t) \equiv x(t) \equiv \dot{x}(t) \equiv u(t)$ (the control u is singular in that interval). We choose the stock as that result suggests: $\Xi = \{P_1, P_2, P_3\}$, $P_1(x) \equiv -1$, $P_2(x) \equiv 0$, $P_3(x) \equiv +1$. P_1 and P_3 are boundary control procedures, and P_2, a candidate singular one. A consequence of the fact that all control procedures are constant is that the adjoint final value problem of the maximum principle is identical with the induced adjoint problem (3).

Our aim is to show the convergence consequences of different choices for the parameter γ in Step 2. Assume that the rule, used to determine the inserted control procedure P and the insertion point θ, is as follows

$$D_{P,\theta}(\xi^k) = \max_{p \in \Xi} \min_t \{D_{p,t}(\xi^k) : D_{p,t}(\xi^k) \le \gamma D(\xi^k)\}, \quad k = 1, 2, \ldots. \tag{16}$$

Let the starting decision be $(P_1, (0, T))$, equal to ξ^1 because the sequence $(0, T)$ is stationary. In the corresponding induced solution, $x(t) = 1 - t$ and $\psi(t) = t - \frac{1}{2}t^2$. The cost value $\Sigma(\xi^1)$ is equal to $\frac{1}{3}$. We calculate $D_{P_1,\,t}(\xi^1) = 0$, $D_{P_2,\,t}(\xi^1) = -\psi(t)$ and $D_{P_3,\,t}(\xi^1) = -2\psi(t)$, for $t \in [0, T]$. The decision ξ^1 is not stationary, since $D(\xi^1) = D_{P_3,\,1}(\xi^1) = -1$. Suppose first that $\gamma = 0.4$. We choose $P = P_2$ and $\theta = 1$ in accordance with (16). The spike generation in Step 2 yields the decision $\bar{\xi} = (\bar{S}, \bar{\tau})$, where $\bar{S} = (P_1, P_2, P_1)$ and $\bar{\tau} = (0,\ 1,\ 1,\ T)$. The subsequent gradient optimization (Step 1 of the algorithm) leads to a stationary decision ξ^2 with the induced solution $u(t) = -1$, $x(t) = 1 - t$, $\psi(t) = -\frac{1}{2}(t-1)^2$ for $t < 1$, and $u(t) = x(t) = \psi(t) = 0$ for $t \ge 1$. It is easy to verify that this solution is optimal in the original problem, with the cost $Q(u) = \Sigma(\xi^2) = \frac{1}{6}$.

We shall now show that with the value of γ equal to 1, the MSE algorithm becomes infinite. The rule (16) then takes the form $D_{P,\,\theta}(\xi^k) = D(\xi^k)$. In compliance with this rule, the spike generation in every iteration inserts the control procedure $P = P_3$ at $\theta = 1$. Successively, the decisions $\xi^k = (S^k, \tau^k)$, $k = 1, 2, \ldots$, are produced, with $N(\xi^k) = 2k - 1$ and

$$S_i^k = (-1)^i, \quad \tau_i^k = 1 + \frac{2i - 1}{4k - 3}, \text{ for } i = 1, \ldots, N(\xi^k).$$

Simple calculations give the values of $D(\xi^k)$ and $\Sigma(\xi^k)$ for an arbitrary k

$$D(\xi^k) = -\frac{1}{(4k-3)^2}, \quad \Sigma(\xi^k) = \frac{1}{6}\left(1 + \frac{1}{(4k-3)^2}\right).$$

Figure 5 presents the control obtained in iteration 6 (induced by ξ^6), and the plots of the negative parts of $D_{P_1,\,t}(\xi^6)$, $D_{P_2,\,t}(\xi^6)$ and $D_{P_3,\,t}(\xi^6)$.

Example 4. In this example, the stock Ξ contains enough control procedures to construct the optimal control structure, which is finite. Still, the MSE algorithms may exhibit infinite convergence for any choice of the parameter γ, producing infinite sequences of nonstationary decisions $\xi^k = (S^k, \tau^k)$, $k = 1, 2, \ldots$,

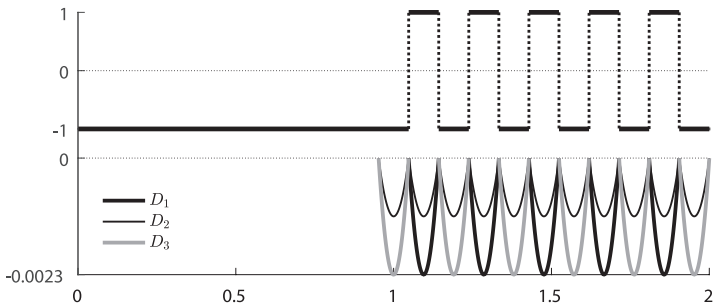

Figure 5. The control induced by ξ^6 is plotted in the upper part; the negative arcs of the functions $D_i(t) = D_{P_i,t}(\xi^6)$ are plotted below (time on horizontal axes)

such that every point τ^k is stationary, and the sequence of cost values $\Sigma(\xi^k)$ is strictly decreasing. This example is a modification of Example 3, where we substitute x_1 for the state variable x and introduce another state variable x_2 to represent time, satisfying $\dot{x}_2(t) = 1$, $t \in [0, T]$, $x_2(0) = 0$. The cost functional and the control constraints are not changed, neither are the optimality conditions, the adjoint final value problem, nor the optimal control. The essential new feature is in the construction of Ξ. Define the auxiliary functions, for $t < T$

$$\begin{aligned}
d_1(t) &= a_1(t-T)^7 s(t),\\
d_2(t) &= \dot{d_1}(t) = 7a_1(t-T)^6\, s(t) - a_1a_2(t-T)^5\, c(t),\\
d_3(t) &= \dot{d_2}(t) = a_1(t-T)^3(42(t-T)^2 - a_2^2)\, s(t) - 12a_1a_2(t-T)^4\, c(t),\\
d_4(t) &= \dot{d_3}(t) = a_1(t-T)^2(210(t-T)^2 - 15a_2^2)\, s(t)\\
&\qquad + a_1a_2(t-T)(a_2^2 - 90(t-T)^2)\, c(t),
\end{aligned}$$

where $s(t) = \sin(a_2(t-T)^{-1})$, $c(t) = \cos(a_2(t-T)^{-1})$. For $t \geq T$, $d_i(t) = 0$, $i = 1, ..., 4$. The parameters a_1 and a_2 are chosen so that $d_3(1) = -1$, $d_4(1) = 0$. Hence

$$a_1 = \frac{1}{12a_2\cos a_2 + (a_2^2 - 42)\sin a_2}, \tag{17}$$

and a_2 is an arbitrary nonzero solution of the equation

$$\tan a_2 = \frac{a_2(a_2^2 - 90)}{15(a_2^2 - 14)}. \tag{18}$$

Let the stock consist of only two procedures, $\Xi = \{P_1, P_2\}$, with

$$P_1(x) = \left\{ \begin{array}{ll} -1, & x_2 \leq 1 \\ d_3(x_2), & x_2 \geq 1 \end{array} \right\}, \quad P_2(x) \equiv 0.$$

P_1 and P_2 are continuously differentiable. The optimal solution (see Example 3) is induced by the decision $\xi = (S, \tau)$, $S = (P_1, P_2)$, $\tau = (0,\ 1,\ T)$.

Note that for the stock consisting of two control procedures, the rules of choice of P and θ, proposed in Examples 1 and 3, simplify to the form $D_{P,\theta}(\xi) = D(\xi)$. In consequence, the MSE algorithms with those generation rules do not depend on γ. Consider now a decision $\xi = (S, \tau)$, $S = (P_1, P_2, P_1)$, $\tau = (0,\ \tau_1, \tau_2, T)$, with $\tau_1 \le 1$ and $\tau_2 > 1$. The corresponding induced control and trajectory x_1 are expressed by

$$u(t) = \left\{ \begin{array}{ll} -1, & t < \tau_1 \\ 0,\ \tau_1 \le t < \tau_2 \\ d_3(t), & t \ge \tau_2 \end{array} \right\},\ x_1(t) = \left\{ \begin{array}{ll} 1-t, & t \le \tau_1 \\ 1-\tau_1, & \tau_1 \le t \le \tau_2 \\ 1-\tau_1+d_2(t)-d_2(\tau_2), & t \ge \tau_2 \end{array} \right\}.$$

The adjoint trajectory satisfies $\psi(t) = \int_T^t x_1(s)\mathrm{d}s$, and so

$$\psi(t) = \left\{ \begin{array}{ll} \psi(\tau_1) + t - \tau_1 - \frac{1}{2}(t^2 - \tau_1^2), & t \le \tau_1 \\ \psi(\tau_2) + (1-\tau_1)(t-\tau_2), & \tau_1 \le t \le \tau_2 \\ (1-\tau_1-d_2(\tau_2))(t-T) + d_1(t), & t \ge \tau_2 \end{array} \right\}.$$

To examine the stationarity conditions of the structural nodes, calculate

$$\frac{\partial \Sigma}{\partial \tau_1} = \psi(\tau_1)(P_2(x(\tau_1)) - P_1(x(\tau_1)) = \psi(\tau_1),$$

$$\frac{\partial \Sigma}{\partial \tau_2} = \psi(\tau_2)(P_1(x(\tau_2)) - P_2(x(\tau_2)) = \psi(\tau_2) P_1(x(\tau_2)).$$

Hence, the conditions take the form $\psi(\tau_1) = 0$ and $\psi(\tau_2)\, d_3(\tau_2) = 0$. For that set of equations, we compute solutions with $\tau_1 = 1$. The value of $\tau_2 > 1$ thus satisfies $\psi(\tau_2) = 0$, or $d_2(\tau_2)(\tau_2 - T) - d_1(\tau_2) = 0$, whence

$$\tan\left(\frac{a_2}{T-\tau_2}\right) = \frac{a_2}{6(T-\tau_2)}. \tag{19}$$

Evidently, equation (19) has in $]1, T[$ an infinite number of solutions with respect to τ_2. They can be arranged in a strictly increasing sequence τ_2^i, $i = 1, 2, ...$, convergent to T. Let $\xi^i = (S, \tau^i)$, $\tau^i = (0,\ 1, \tau_2^i, T)$. The sequence of cost values

$$\Sigma(\xi^i) = \tfrac{1}{6} + \tfrac{1}{2} \int_{\tau_2^i}^{T} (d_2(t) - d_2(\tau_2^i))^2 \mathrm{d}t$$

strictly decreases and tends to the optimal value of $\frac{1}{6}$.

Let us now discuss Algorithm 1 with the generation rule $D_{P,\theta}(\xi) = D(\xi)$ and with a steepest descent gradient algorithm, such that the linesearch returns the nearest stationary point on the steepest descent halfline. Let the starting decision be ξ^1, defined above. It is easy to check that the algorithm then produces an infinite sequence of decisions $\xi^k = (S^k, \tau^k)$, $k = 1, 2, ...$, with $S^k = (P_1, P_2, P_1)$ and $\tau^k = (0,\ 1, \tau_2^k, T)$, τ_2^k being the kth term of the sequence of solutions of (19). Attempts to achieve finite convergence by replacing the rule of choice of P and θ by a rule depending on γ, e.g., $D_{P,\theta}(\xi) = \gamma D(\xi)$, prove futile, yielding more complex structures S^k.

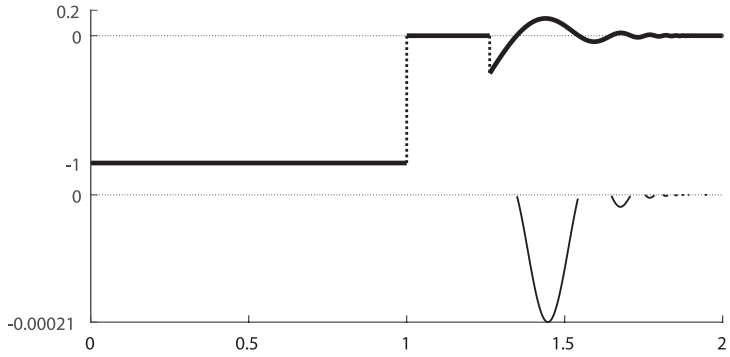

Figure 6. The control induced by ξ^1 (top) and the negative part of $D_{P_2,t}(\xi^1)$ (bottom) (time on horizontal axes)

For $a_1 = 0.0218175208363471$ and $a_2 = 5.276866884891947$ taken as the solutions of (17) and (18), $\tau_2^1 \cong 1.262620582159035$ is the smallest solution of (19), greater than 1. Figure 6 shows the control u induced by the decision ξ^1, and the negative part of the slope index $D_{P_2,\,t}(\xi^1)$. To end the discussion of this example, note that Algorithm 1 with Step 1 complete in every iteration, would obviously be finitely convergent to the optimal solution.

7. Observations on finite convergence

Provided a reasonable choice of the stock Ξ and sufficient computational accuracy, the MSE algorithms exhibit a certain desirable feature. In the optimal control problems, where the conditions of Pontryagin's maximum principle are only fulfilled by controls with finite structures, the MSE method usually finds a solution, satisfying the maximum principle, after a finite number of iterations (as in Example 1). We cannot offer an in-depth theoretical explanation of that phenomenon, instead, we shall merely give a few observations.

To start with, consider a 'complete search' optimization method, which explains what can be aimed at in the construction of MSE algorithms. Let σ be a sequence of all elements of Ξ, and let σ_k denote its k-fold concatenation. The complete search algorithm is as follows.

CS Algorithm.

Step 0. Set $k := 1$.

Step 1. Find $\xi^k = (S^k, \tau^k)$, such that $\Sigma(\xi^k) = \min\{\Sigma(\xi) : \xi = (\sigma_k, \tau), \tau \in \Theta_{N(\xi)}\}$, τ^k is strictly increasing and S^k is a subsequence of σ_k.

Step 2. If $D(\xi^k) = 0$, stop. Otherwise, set $k := k + 1$, and return to Step 1.

Note that for every decision $\xi = (S, \tau)$ of length shorter than k, there is

an equivalent decision $\bar{\xi} = (\sigma_{k-1}, \bar{\tau})$ with some $\bar{\tau} \in \Theta_{N(\bar{\xi})}$, where S and τ are subsequences of σ_{k-1} and $\bar{\tau}$, respectively (see Lemma 5). Hence, it follows that if an optimal decision (of finite length) exists, then the CS algorithm is finite. However, this algorithm can hardly be regarded as a version of the MSE, since Step 1 of iteration $k+1$ does not generally begin with a decision equivalent to ξ^k, and the return to Step 1 after Step 2 brings on a temporary increase in the cost Σ. This is in contrast to Algorithms 1 and 2. Besides, the CS algorithm is impractical for two more reasons. First, the 'multiple generations' in Step 2 produce unnecessarily complex control structures, which results in inefficient optimization, and second, the requirement that Step 1 be complete (in the sense of Section 4) may be difficult to fulfill. That is why neither of these 'complete search' properties has been included in Algorithms 1 and 2 (the completeness of Step 1, however, is assumed in most of the results below).

We now give several observations on finite convergence of Algorithm 1. The trivial case of $|\Xi| = 1$ will be left out.

Observation 1. *Let Step 1 of Algorithm 1 be complete for every k. Then, Algorithm 1 is finite (that is, $D(\xi^k) = 0$ for some k) if and only if there is a constant M, such that $N(\xi^k) \leq M$ for every k.*

Proof. This is straightforward from the fact that the sequence $\Sigma(\xi^k)$, $k = 1, 2, \ldots$, strictly decreases, and the number of control structures with a given length is finite. ■

The next observation is an obvious consequence.

Observation 2. *Algorithm 1 is finite, if Step 1 is complete for every k, $|\,\Xi\,| = 2$, and there is an optimal decision.*

Observation 3. *Algorithm 1 is finite, if Step 1 is complete for every k, all control procedures in Ξ are constant, and the right-hand side of the state equation (1) is linear of the form $f(x,u) = Ax + u$, where A is a constant $n \times n$ matrix.*

Proof. Let $P_1, P_2 \in \Xi$, and let ψ satisfy $\dot{\psi}(t) = -A^\top \psi(t)$, $t \in [0,T]$, $\psi(T) = a$. It follows from the theory of linear differential equations with constant coefficients that if the function $t \mapsto \psi(t)^\top (P_1 - P_2)$ is not identically zero, then the number of its roots is less than some integer M_0, independent of P_1, P_2, and $a \in R^n$. Now, it is enough to notice that $M = |\,\Xi\,|^2 M_0$ is the constant in Observation 1. ■

Observation 4. *Let B be the compact set defined in Assumption 1. Moreover, assume that*

(i) Step 1 is complete for every k,

(ii) all control procedures in Ξ are constant,

(iii) the right-hand side of the state equation (1) has the form $f(x,u) = f^0(x) + f^1(x)\,u$ with $n = 2$ and $m = 1$ (that is, $x(t) \in R^2$ and $u(t) \in R$),

(iv) the function $f^2(x) = \partial f^1(x)^\top f^0(x) - \partial f^0(x)^\top f^1(x)$ is Lipschitz continuous in B,

(v) $\partial q(x) \neq 0 \ \ \forall x \in B$,
(vi) $\det [f^1(x) \ \ f^2(x)] \neq 0 \ \ \forall x \in B$.
Algorithm 1 is then finite.

PROOF. Let $\tau^k_{i_k} - \tau^k_{i_k-1} = \min\{\tau^k_i - \tau^k_{i-1} : \ \ i \in 1, ..., N(\xi^k)\}$, $k = 1, 2, ...$ Suppose, contrary to the claim, that Algorithm 1 is infinite. Then by Observation 1, $\tau^k_{i_k} - \tau^k_{i_k-1} \to 0$ as $k \to \infty$. From (ii) and the stationarity of τ^k (see (5)) it follows that

$$\phi(x^k(\tau^k_{i_k-1}), \psi^k(\tau^k_{i_k-1})) = \phi(x^k(\tau^k_{i_k}), \psi^k(\tau^k_{i_k})) = 0, \ \ k = 1, 2, ..., \tag{20}$$

where $\phi(x, \psi) = \psi^\top f^1(x)$, and x^k and ψ^k denote, respectively, the solutions of (2) and (3) in iteration k. The sequence $x^k(\tau^k_{i_k})$, $k = 1, 2, ...$, is contained in B, and it follows from Lemma 2 that all ψ^k are uniformly bounded. Thus, the sequence of pairs $(x^k(\tau^k_{i_k}), \psi^k(\tau^k_{i_k}))$, $k = 1, 2, ...$, has an accumulation point $(\bar{x}, \bar{\psi})$ with $\bar{x} \in B$. Denote $\dot{\phi}(x, \psi) = \psi^\top f^2(x)$. The function $t \mapsto \phi(x^k(t), \psi^k(t))$ is continuously differentiable and by (iv), its derivative $t \mapsto \dot{\phi}(x^k(t), \psi^k(t))$ is Lipschitz continuous with a Lipschitz constant independent of k. Hence

$$\dot{\phi}(x^k(\tau^k_{i_k}), \psi^k(\tau^k_{i_k})) \to 0 \text{ as } k \to \infty. \tag{21}$$

From (20) and (21), $\phi(\bar{x}, \bar{\psi}) = 0$ and $\dot{\phi}(\bar{x}, \bar{\psi}) = 0$. By virtue of (v) there is an $\varepsilon > 0$, such that $|| \psi^k(\tau^k_{i_k}) || \geq \varepsilon$ for every k, whence $\bar{\psi} \neq 0$. We conclude that $\det [f^1(\bar{x}) \ \ f^2(\bar{x})] = 0$, which contradicts the assumption (vi). ■

The next observation is about finite convergence of Algorithm 2.

OBSERVATION 5. *Assume the following:*
(i) all control procedures in Ξ *are constant,*
(ii) the right-hand side of the state equation (1) has the form $f(x, u) = f^0(x) + bu$, *where* $f^0(x) = \mathrm{col}(f^0_1(x_1), ..., f^0_n(x_n))$, $f^0_j : R \to R$, $j = 1, ..., n$, *and* $b = \mathrm{col}(b_1, ..., b_n) \in R^n$,
(iii) for every $x \in B$, $\partial q(x)^\top b \neq 0$ *and* $(\partial q(x)^\top b)\, \partial_j q(x)\, b_j \geq 0$, $j = 1, ..., n$, *where* B *is defined in Assumption 1 and* $\partial_j q(x)$ *denotes the* j*th component of* $\partial q(x)$.
Then, Algorithm 2 is finite.

PROOF. Assume that in iteration k of the algorithm, for $k = 1, 2, ...$, θ_k is the point at which the spike is inserted in Step 2, and x^k, ψ^k denote, respectively, the solutions of (2) and (3). Suppose that the algorithm is infinite, and so $D(\xi^k) < 0$ for every k and $D(\xi^k) \to 0$ as $k \to \infty$. Therefore, there exist $P_1, P_2 \in \Xi$ and an infinite, strictly increasing sequence of positive integers k_i, $i = 1, 2, ...$, such that $D(\xi^{k_i}) = \psi^{k_i}(\theta_{k_i})^\top b(P_1 - P_2)$. Let

$$w_i(t) = \psi^{k_i}(t)^\top b = -\partial q(x^{k_i}(T))^\top \mathrm{diag}(e_{i1}(t), ..., e_{in}(t))\, b, \ \ t \in [0, T],$$
$$e_{ij}(t) = \exp\left(\int_t^T \partial f^0_j(x^{k_i}_j(s))\, \mathrm{d}s \right), \ \ j = 1, ..., n.$$

By virtue of the assumptions on (1) in Section 1, Assumption 1, the assumption (iii) and the Weierstrass theorem, there is an $\eta > 0$, such that $|\,\partial q(x)^\top b\,| \geq \eta$ for every $x \in B$. Similarly, it can be shown that there is a positive real number e, such that $e_{ij}(t) \geq e > 0$ for every positive integer i, every $j \in \{1, ..., n\}$ and every $t \in [0, T]$. Hence,

$$|\,w_i(t)\,| = \sum_{j=1}^{n} e_{ij}(t)\,|\partial_j q(x^{k_i}(T))\,b_j| \geq e\eta > 0$$

for every $t \in [0, T]$ and every i. Thus, we have arrived at a contradiction. ■

Note that an analogous observation is valid for Algorithm 1, if additionally, its Step 1 is complete for every k.

Observation 6. *If the system (1) is of order 1 ($n = 1$) and all functions f_P, $P \in \Xi$, are pairwise different in the whole set B, determined in Assumption 1, then Algorithms 1 and 2 are finite (they stop after a number of iterations not greater than the number of elements in Ξ).*

Proof. Let x and ψ be the solutions of (2) and (3), respectively, corresponding to a decision (S, τ). A structural node $\tau_i \in]0, T[$ is stationary, if $\psi(\tau_i)(f_{S_{i+1}}(x(\tau_i)) - f_{S_i}(x(\tau_i))) = 0$, or $\psi(\tau_i) = 0$. Consequently, the control structure in every iteration of the algorithm consists of only one term and/or the adjoint function ψ is identically zero. Since the sequence of cost values is strictly decreasing, the claim follows. ■

8. Stationary decisions and the maximum principle

We finish the paper with a discussion of relationships between the MSE stationarity conditions and Pontryagin's maximum principle, which will allow a better understanding of the obtained results on convergence. One of the basic premises behind the MSE is that the optimization algorithm should naturally stop only when the current optimal control approximation satisfies the necessary optimality condition of the maximum principle. In other words, we wish that the controls induced by stationary decisions be always extremal. This requirement is easy to accomplish if the stock contains appropriately parameterized control procedures, as described in Section 2. Here we shall show that the equivalence between the decision stationarity and the induced control, being extremal, can be sometimes proved also for procedures depending merely on state, considered in this article.

Define the Pontryagin function $H : R^n \times R^n \times R^m \to R$, $H(\psi, x, u) = \psi^\top f(x, u)$, and the set of all control values attainable at an arbitrary point $x \in R^n$, $U_x = \{v \in U : \exists\, P \in \Xi,\ v = P(x)\}$. Let u_ξ be the control induced by a node-stationary decision ξ, and let x_ξ and ψ_ξ denote, respectively, the corresponding solutions of (1) and (3). Note first that a certain weak maximum principle holds at every stationary point in the decision space.

Observation 7. *The stationarity condition of the decision ξ, $D(\xi) = 0$, is equivalent to the following maximum condition*

$$H(\psi_\xi(t), x_\xi(t), u_\xi(t)) \geq H(\psi_\xi(t), x_\xi(t), v) \ \ \forall v \in U_{x_\xi(t)} \ \forall t \in [0, T].$$

This observation is straightforward from the definition $D(\xi) = \min\{D_{P,\theta}(\xi) : P \in \Xi,\ 0 \leq \theta \leq T\}$. Thus, $\psi_\xi(t)^\top (f(x_\xi(t), u_\xi(t)) - f(x_\xi(t), P(x_\xi(t)))) \geq 0$ for every $P \in \Xi$ and every $t \in [0, T]$.

The control u_ξ is *extremal*, if it fulfills the necessary optimality condition of the classical Pontryagin's maximum principle for the optimal control problem of Section 1, that is, if the relationship

$$H(\psi_\xi(t), x_\xi(t), u_\xi(t)) \geq H(\psi_\xi(t), x_\xi(t), v) \ \ \forall v \in U \text{ a.e. } t \in [0, T],$$

holds together with

$$\partial_2 F(t, x_\xi(t))\, \psi_\xi(t) = \partial_2 H(\psi_\xi(t), x_\xi(t), u_\xi(t)) \text{ a.e. } t \in [0, T], \tag{22}$$

in (3). It is obvious that if the control u_ξ is extremal, then $D(\xi) = 0$. Let us ask: when the decision stationarity is sufficient for the induced control to be extremal? The following is evident from the definition.

Observation 8. *Assume that*
(i) U is finite, $U = \{e^1, ..., e^k\}$,
(ii) $\Xi = \{P_1, ..., P_k\}$ and $P_i(x) \equiv e^i$, $i = 1, ..., k$,
(iii) ξ is stationary.
Then, the control u_ξ is extremal.

Observation 9. *Assume that*
(i) there is a finite set E, such that $E \subset U \subset \mathrm{Conv}(E)$,
(ii) for every $e \in E$ there is a $P \in \Xi$, such that $P(x) \equiv e$,
(iii) the right-hand side of (1) has the form $f(x, u) = f^0(x) + f^1(x)\, u$,
(iv) equality (22) is true,
(v) ξ is stationary.
Then, the control u_ξ is extremal.

Proof. By (i), for every $v \in U$ there exist $a_i \geq 0$ and $e^i \in E$, $i = 1, \ldots, k$, such that $\sum_{i=1}^k a_i = 1$ and $v = \sum_{i=1}^k a_i e^i$. From (ii), (iii) and (v), we have $\psi_\xi(t)^\top f^1(x_\xi(t))(u_\xi(t) - e^i) \geq 0$ for $i = 1, ..., k$, and for every $t \in [0, T[$. Hence $\sum_{i=1}^k a_i \psi_\xi(t)^\top f^1(x_\xi(t))(u_\xi(t) - e^i) \geq 0$, and $\psi_\xi(t)^\top f^1(x_\xi(t))(u_\xi(t) - v) \geq 0$ $\forall v \in U\ \forall t \in [0, T[$. ∎

In the last two observations we make the following assumption.

Assumption 3. *Let $\xi = (S, \tau)$ and let S_{ij} denote the jth component of the vector S_i, $i = 1, ..., N(\xi)$, $j = 1, ..., m$. Assume also that*
(i) $U = [u_{1\,\min}, u_{1\,\max}] \times [u_{2\,\min}, u_{2\,\max}] \times \ ... \ \times [u_{m\,\min}, u_{m\,\max}]$,
(ii) S_i, $i = 1, ..., N(\xi)$, are continuously differentiable,

(iii) for $i = 1, ..., N(\xi)$, $j = 1, ..., m$, either $u_{j\min} < S_{ij}(x_\xi(t)) < u_{j\max}$ for a.e. $t \in [\tau_{i-1}, \tau_i]$, or S_{ij} is constant and $S_{ij}(x) \in \{u_{j\min}, u_{j\max}\}$ for every $x \in R^n$,
(iv) ξ is stationary.

The next observation explains, why the control obtained in Example 1 satisfies the optimality condition of the maximum principle.

Observation 10. *Let Assumption 3 hold. Assume also that*
(i) for every vertex e of U there is a $P \in \Xi$, such that $P(x) \equiv e$,
(ii) the right-hand side of (1) has the form $f(x, u) = f^0(x) + f^1(x)\, u$.
Then, the control u_ξ is extremal.

Proof. By definition, $D_{P,t}(\xi) = \psi_\xi(t)^\top f^1(x_\xi(t))(u_\xi(t) - P(x_\xi(t)))$. Denote $u = u_\xi(t)$, $P = P(x_\xi(t))$, $\phi = f^1(x_\xi(t))^\top \psi_\xi(t)$. Then

$$D_{P,t}(\xi) = \phi^\top (u - P) = \sum_{j=1}^{m} \phi_j (u_j - P_j).$$

Using (iv) in Assumption 3 and an argument similar to that used for Observation 9 we can prove that $\phi_j(u_j - v) \geq 0$ for every $v \in [u_{j\min}, u_{j\max}]$ and $j = 1, ..., m$. This implies $\phi_j = 0$, if $u_{j\min} < u_j < u_{j\max}$. Consider now the right-hand side of (3). For $t \in [\tau_{i-1}, \tau_i[$ and $i = 1, ..., N(\xi)$

$$\partial_2 F(t, x_\xi)\, \psi_\xi = \partial_1 f(x_\xi, u_\xi)\, \psi_\xi + \partial S_i(x_\xi)\, f^1(x_\xi)^\top \psi_\xi.$$

Fix i and let σ_j denote the jth column of $\partial S_i(x_\xi(t))$, $j = 1, ..., m$. Then

$$\partial S_i(x_\xi)\, f^1(x_\xi)^\top \psi_\xi = \sum_{j=1}^{m} \sigma_j \phi_j.$$

If $u_{j\min} < S_{ij}(x_\xi(t)) < u_{j\max}$ for a.e. $t \in [\tau_{i-1}, \tau_i]$, then $\phi_j = 0$. If S_{ij} is constant, then $\sigma_j = 0$. Thus, by Assumption 3 (iii), $\partial S_i(x_\xi)\, f^1(x_\xi)^\top \psi_\xi \equiv 0$ and (22) holds. The claim follows from Observation 9. ∎

Part (i) of Assumption 3 and the assumption (ii) of Observation 10 are obviously fulfilled in Example 1. To satisfy the assumption (i) of Observation 10, we have to complement the stock Ξ with a procedure $P_6 = \mathrm{col}(u_{1m}, u_{2m})$. This, however, affects neither the process of solving the problem nor its result, because pressing the gas and brake pedals at the same time can be excluded from the search for optimal control. The remaining parts of Assumption 3 are true for the final decision ξ^4. Note that due to (22), the solutions of the final value problems (3) and (13) coincide.

In the last observation we break free from the assumption that the state equation is affine in control – at the cost of additional requirements with respect to the stock and regularity.

OBSERVATION 11. *Let Assumption 3 be true. Let also for* $j = 1, ..., m$, *and* $t \in [0, T]$

$$H_j(v,t) = H(\psi_\xi(t), x_\xi(t), u_{\xi 1}(t), ..., u_{\xi,j-1}(t), v, u_{\xi,j+1}(t), ..., u_{\xi m}(t)).$$

Assume that

(i) $\Xi = \Xi_1 \times \ ... \ \times \Xi_m$, *where* Ξ_j, $j = 1, \cdots, m$, *is a finite set of appropriately regular functions* $R^n \to [u_{j\min}, u_{j\max}]$*; for every j,* Ξ_j *contains procedures* $P_{j\min}(x) \equiv u_{j\min}$ *and* $P_{j\max}(x) \equiv u_{j\max}$,

(ii) f in (1) is continuously differentiable in its both arguments,

(iii) $\partial_1 H_j(u_{\xi j}(t), t) = 0$, $t \in [\tau_{i-1}, \tau_i[$, *for every pair i, j, such that* $u_{j\min} < S_{ij}(x_\xi(t)) < u_{j\max}$ *for a.e.* $t \in [\tau_{i-1}, \tau_i]$,

(iv) If $\partial_1 H_j(v,t) = 0$ *for some* $v \in [u_{j\min}, u_{j\max}]$ *and* $t \in [0, T]$, *then there is a* $P \in \Xi_j$, *such that* $P(x_\xi(t)) = v$.

Then, the control u_ξ *is extremal.*

PROOF. Let J_i be the set of all $j \in \{1, ..., m\}$, such that $u_{j\min} < S_{ij}(x_\xi(t)) < u_{j\max}$ for a.e. $t \in [\tau_{i-1}, \tau_i]$. By Assumption 3 (iii), the right-hand side of (3) can be written in the form, for $t \in [\tau_{i-1}, \tau_i[$ and $i = 1, ..., l(S)$,

$$\partial_2 F(t, x_\xi)\psi_\xi = \partial_2 H(\psi_\xi, x_\xi, u_\xi) + \sum_{j \in J_i} \sigma_{ij} \partial_1 H_j(u_{\xi j}(t), t),$$

where σ_{ij} denotes the jth column of $\partial S_i(x_\xi(t))$. It then follows from (iv) that (22) is true. Now, it is straightforward from (i) and Assumption 3 (iv) that for every t and every j

$$H_j(u_{\xi j}(t), t) \geq H_j(P(x_\xi(t)), t) \ \ \forall P \in \Xi_j.$$

We thus have, by (i), Assumption 3 (iv), Assumption 3 (iii), and (iv)

$$H_j(u_{\xi j}(t), t) \geq H_j(v, t) \ \ \forall v \in [u_{j\min}, u_{j\max}],$$

whence the claim. ■

Acknowledgements

The authors thank the two anonymous reviewers for their valuable comments and suggestions. Thanks are also due to the editors for their suggestion to extend the paper.

References

AXELSSON, H., WARDI, Y., EGERSTEDT, M. and VERRIEST, E.I. (2008) Gradient descent approach to optimal mode scheduling in hybrid dynamical systems. *Journal of Optimization Theory and Applications*, **136** (2), 167–186.

Korytowski, A. and Szymkat, M. (2010) Consistent Control Procedures in the Monotone Structural Evolution. Part 1: Theory. In: M. Diehl et al., *Recent Advances in Optimization and its Applications in Engineering*, Springer, 247–256.

Osmolovskii, N.P. and Maurer, H. (2012) Applications to Regular and Bang–Bang Control: Second-Order Necessary and Sufficient Optimality Conditions in Calculus of Variations and Optimal Control. *SIAM Advances in Design and Control.* **DC 24**, SIAM Publications, Philadelphia.

Schättler, H. and Ledzewicz, U. (2012) *Geometric Optimal Control. Theory, Methods and Examples.* Springer.

Sirisena, H.R. (1974) A gradient method for computing optimal bang-bang control. *International Journal of Control*, **19**, 257-264.

Szymkat, M. and Korytowski, A. (2003) Method of monotone structural evolution for control and state constrained optimal control problems. *European Control Conference ECC 2003*, University of Cambridge, U.K., September 1-4.

Szymkat, M. and Korytowski, A. (2007) Evolution of structure for direct control optimization. *Discussiones Mathematicae. Differential Inclusions, Control and Optimization*, **27**, 165–193.