J Glob Optim (2009) 44:213–225
DOI 10.1007/s10898-008-9319-5

A filled function method for optimal discrete-valued control problems

C. Z. Wu · K. L. Teo · V. Rehbock

Received: 16 March 2007 / Accepted: 15 June 2008 / Published online: 16 July 2008

Abstract In this paper, we present a new approach to solve a class of optimal discrete-valued control problems. This type of problem is first transformed into an equivalent two-level optimization problem involving a combination of a discrete optimization problem and a standard optimal control problem. The standard optimal control problem can be solved by existing optimal control software packages such as MISER 3.2. For the discrete optimization problem, a discrete filled function method is developed to solve it. A numerical example is solved to illustrate the efficiency of our method.

Keywords Discrete-valued control · Optimal control · Parametrization · Filled function

1 Introduction

Optimal control problems arise in a variety of fields, such as engineering, economics, and biomedicine. However, in many practical applications, the control can only take values from a discrete set, such as switched amplifier designs [7], optimal driving strategies for trains [2] and the management of batteries in a submarine [6]. For these optimal control problems, we need to find switching points and the corresponding control values from a discrete set. Since the control evolves in a discrete set and the switching points are continuous variables, these optimal control problems are mixed integer optimization problems. So far, there are no efficient algorithms with polynomial-time complexity for solving these problems. They are in fact, NP hard.

In [5], a method is developed for solving this class of optimal discrete-valued control problems. A time scaling transformation is used to transform the optimal discrete-valued

C. Z. Wu (✉)
Department of Mathematics, Chongqing Normal University, Chongqing, People’s Republic of China
e-mail: changzhiwu@yahoo.com

K. L. Teo · V. Rehbock
Department of Mathematics and Statistics, Curtin University of Technology, Perth, Western Australia, Australia

control problem into an optimal parameter selection problem which is solvable by existing optimal control techniques. However, if the maximum number of switchings is pre-fixed, the solution obtained by the method in [5] may not give a feasible solution to the original problem, as the transformed problem in [5] is not equivalent to the original under this condition. In this paper, we propose a two-level optimization approach to solve this optimal discrete-valued control problem. In the first level, the time scaling transformation used in [4] and [5], is introduced to transform the switching points for a given switching sequence into pre-fixed knots in a new time horizon. The resulting problem is a standard optimal parameter selection problem and hence solvable by existing optimal control methods such as MISER [3]. In the second level, a discrete filled function method developed in [1,11] is used to obtain a method for finding the optimal switching sequence.

The rest of the paper is organized as follows. In Sect. 2, we formulate the problem to be solved. In Sect. 3, we reformulate our problem as a two-level optimization problem. The first level is a standard optimal control problem, and a gradient-based method is introduced and hence the optimal control software package MISER can be used to solve it. The second level is a discrete optimization problem. In Sect. 4, we introduce a discrete filled function method to solve the discrete optimization problem. In Sect. 5, a numerical example is solved using our method. In Sect. 6, we give some concluding remarks.

2 Problem formulation

Consider a process described by the following differential equations defined on $(0, T]$:

$$\dot{\mathbf{x}} = \mathbf{f}\left(\mathbf{x}, \mathbf{u}, t\right), \tag{1}$$

with the initial condition

$$\mathbf{x}\left(0\right) = \mathbf{x}_0, \tag{2}$$

where $\mathbf{x} \in \mathbb{R}^n$ and $\mathbf{u} \in \mathbb{R}^m$ are, respectively, the state and control vectors. T is the terminal time. The function $\mathbf{f}$ is assumed to be continuously differentiable with respect to all its arguments.

Let $\mathbf{U}$ be defined by $\mathbf{U} = \{\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_K\}$. A function $\mathbf{u}$ is said to be an admissible control if

$$\mathbf{u}\left(t\right) = \mathbf{b}_i,\ t \in \left[\tau_i, \tau_{i+1}\right], i = 0, 1, \ldots, N-1, \tag{3}$$

where $\mathbf{b}_i \in \mathbf{U}$, and $\tau_1, \tau_2, \ldots, \tau_{N-1}$ are the switching points of $\mathbf{u}$ satisfying:

$$0 = \tau_0 < \tau_1 < \tau_2 < \cdots < \tau_{N-1} < \tau_N = T.$$

Let $\mathcal{U}$ be the class of all such admissible controls. We assume that $N-1$ is the maximum possible number of switching points for any $u \in \mathcal{U}$. For each $\mathbf{u} \in \mathcal{U}$, we integrate Eq. 1 successively over each interval $\left[\tau_i, \tau_{i+1}\right]$, $i = 0, 1, \ldots, N-1$. The obtained $\mathbf{x}\left(t\right)$ is continuous and piecewise differentiable on $(0, T)$. It is called the state of the system (1)–(2) corresponding to $u \in \mathcal{U}$. We assume that the function $\mathbf{f}$ also satisfies the following condition: There exists a constant M, such that

$$\left\|\mathbf{f}\left(\mathbf{x}, \mathbf{u}, t\right)\right\| \leq M\left(1 + \|x\|\right), \tag{4}$$

for all $(\mathbf{x}, \mathbf{u}, t) \in \mathbb{R}^n \times \mathcal{U} \times [0, T]$, where $\|\cdot\|$ denotes the usual norm of $\mathbb{R}^n$. Now, we formally state our optimal control problem as follows:

Given the dynamical system (1), (2), find a $\mathbf{u} \in \mathcal{U}$ such that the cost functional

$$J_0(\mathbf{u}) = \mathbf{\Phi}_0(\mathbf{x}(T)) + \int_0^T \mathcal{L}_0(\mathbf{x}, \mathbf{u})\, dt \tag{5}$$

is minimized subject to the following constraints.

$$\mathbf{h}_i(\mathbf{x}(t)) \geq 0,\ \forall\, t \in [0, T],\ i = 1, \ldots, L. \tag{6}$$

where $\mathbf{\Phi}_0, \mathcal{L}_0, \mathbf{h}_i, i = 1, \ldots, L$, are continuously differentiable functions in their respective arguments. Let this problem be referred to as Problem (P).

A control $\mathbf{u} \in \mathcal{U}$ is said to be a feasible control if it satisfies (6). Let $\mathcal{F}$ be the set of all such feasible controls. We assume that $\mathcal{F}$ is not empty.

Remark 2.1 In solving Problem (P), we need to determine the switching sequence $\mathbf{b}_i, i = 0, 1, \ldots, N-1$, and the switching points $\tau_i, i = 1, \ldots, N-1$. However, the gradients of the cost functional (5) with respect to switching points τ_i are not continuous (see Theorem 5.3.1 of [9]). To overcome this difficulty, we will adopt an existing time scaling transformation to map the switching points into fixed points in a new time horizon. For dealing with the functional inequality constraints (6), we employ the constraint transcription technique developed in Chapter 8 of [9]. For determining the optimal switching sequence, a discrete filled function method is developed in Sect. 4.

3 Problem transformation

Consider Problem (P). For each $i = 1, 2, \ldots, K$, introduce a transformation

$$\mathbf{y}_i = \sum_{j=1}^{K} \mathbf{u}_j v_{i,j} \tag{7}$$

with the following constraints imposed on $v_{i,j}$:

$$v_{i,j} = 0 \text{ or } 1,\ j = 1, 2, \ldots, K,\ i = 1, \ldots, N, \tag{8}$$

$$\sum_{j=1}^{K} v_{i,j} = 1, i = 1, \ldots, N. \tag{9}$$

Let (3) be written as:

$$\mathbf{u}(t) = \mathbf{y}_i,\ t \in \left[\tau_i, \tau_{i+1}\right], i = 0, 1, \ldots, N-1. \tag{10}$$

Also, we introduce the following time scaling transformation

$$\frac{dt}{ds} = \nu(s), \tag{11}$$

with initial condition

$$t(0) = 0, \tag{12}$$

where

$$\nu(s) = \sum_{i=1}^{N} \delta_i \chi_{[i-1,i]}(s), \tag{13}$$

and

$$\sum_{i=1}^{N} \delta_i = T,\ \delta_i \geq 0,\ \ i = 1, \ldots, N. \tag{14}$$

Let $\mathbf{v} = \left[v_1^\top, v_2^\top, \ldots, v_N^\top\right]^\top$ and $v_i^\top = \left[v_{i,1}, v_{i,2}, \ldots, v_{i,K}\right]^\top$, $i = 1, \ldots, N$. Define

$$\mathbf{\Pi} = \left\{\mathbf{v} \in \mathbb{R}^{NK},\ \mathbf{v} \text{ satisfies (8) and (9)}\right\}, \tag{15}$$

and

$$\Delta = \left\{\delta = [\delta_1, \delta_2, \ldots, \delta_N]^\top \in \mathbb{R}^N,\ \sum_{i=1}^{N} \delta_i = T,\ \delta_i \geq 0,\ i = 1, \ldots, N\right\}. \tag{16}$$

Furthermore, we construct

$$g_\varpi(h) = \begin{cases} h, & \text{if } h \leq -\varpi; \\ -(\varpi - h)^2/4\varpi, & \text{if } -\varpi < h < \varpi; \\ 0, & \text{if } h \geq \varpi. \end{cases} \tag{17}$$

Let the following problem be referred to as Problem $\left(TP_{\gamma,\varpi}\right)$:

Subject to the dynamical system

$$\frac{d\mathbf{x}}{ds} = \nu(s)\,\mathbf{f}(\mathbf{x}(s), \mathbf{u}(s), t(s)), \quad \mathbf{x}(0) = \mathbf{x}_0, \tag{18}$$

find a $(\delta, \mathbf{v}) \in \Delta \times \mathbf{\Pi}$ such that

$$J(\delta, \mathbf{v}) = \mathbf{\Phi}_0(\mathbf{x}(N)) + \int_0^N \mathcal{L}_0(\mathbf{x}(s), \mathbf{u}(s))\,ds + \gamma \sum_{i=1}^{L} \int_0^N g_\varpi(h_i(x(s)))\,ds \tag{19}$$

is minimized, where $\gamma > 0$ and $\varpi > 0$ are adjusted parameters, while

$$\mathbf{u}(s) = \sum_{i=1}^{N} \mathbf{y}_i \chi_{[i-1,i]}(s), \tag{20}$$

and $\mathbf{y}_i$ are defined by (7).

Define

$$\Omega = \{\delta \in \Delta : h_i(x(s)) \geq 0,\ i = 1, \ldots, L,\ \text{ for all } s \in [0, N]\},$$

and

$$\mathring{\Omega} = \{\delta \in \Delta : h_i(x(s)) > 0,\ i = 1, \ldots, L,\ \text{ for all } s \in [0, N]\}.$$

We assume that the following condition is satisfied:

Assumption 3.1 For any $\delta \in \mathbf{\Omega}$, there exists a $\bar{\delta} \in \mathring{\mathbf{\Omega}}$, such that

$$\alpha\bar{\delta} + (1 - \alpha)\,\delta \in \mathring{\mathbf{\Omega}} \text{ for all } \alpha \in (0, 1]. \tag{21}$$

This condition, which was first introduced in [9], is a standard assumption made in many papers on semi-infinite optimization problems. See, for example, [8]. According to (4) and Assumption 3.1, we have the following theorem:

Theorem 3.1 *For any $\varpi > 0$, there exists a $\gamma(\varpi) > 0$, such that for all γ, $\gamma \geq \gamma(\varpi)$, if $(\delta^{*,\gamma}, \mathbf{v}^{*,\gamma}) \in \Delta \times \boldsymbol{\Pi}$ is an optimal solution of Problem $(TP_{\gamma,\varpi})$, then the solution, denoted by $x(\cdot|\delta^{*,\gamma}, \mathbf{v}^{*,\gamma})$ of system (18) satisfies $h_i(x(s|\delta^{*,\gamma}, \mathbf{v}^{*,\gamma})) \geq 0, \forall s \in [0, N], i = 1, \ldots, L$. Let such a γ be denoted as $\hat{\gamma}(\varpi)$. Then, as $\varpi \to 0$, the sequence of the optimal solutions of Problems $(TP_{\hat{\gamma}(\varpi),\varpi})$ converges to the optimal solution of Problem (P).*

Proof The proof is similar to that given for Theorem 3.2 in [10]. □

From Theorem 3.1, we see that the solution of Problem (P) can be obtained via solving a sequence of Problems $\left(TP_{\gamma,\varpi}\right)$ by decreasing the value of ϖ while appropriately increasing the value of γ. Clearly, each Problem $\left(TP_{\gamma,\varpi}\right)$ is a mixed-integer programming problem. We propose to decompose it into a two-level optimization problem as follows:

$$\min_{\mathbf{v}\in\boldsymbol{\Pi}} \bar{J}(\mathbf{v}) \tag{22}$$

where

$$\bar{J}(\mathbf{v}) = \min_{\delta\in\Delta} J(\delta, \mathbf{v}). \tag{23}$$

Let the first level problem be referred to as Problem $\left(TP_{\gamma,\varpi}\right)_1$, while the second level problem be referred to as Problem $\left(TP_{\gamma,\varpi}\right)_2$.

For every fixed $\mathbf{v}$, we need to solve Problem $\left(TP_{\gamma,\varpi}\right)_2$. We note that Problem $\left(TP_{\gamma,\varpi}\right)_2$, for each given ϖ and γ, is a standard optimal parameter selection problem. It is to be solved as detailed in the following algorithm.

Algorithm 3.1 For each given $\mathbf{v} \in \boldsymbol{\Pi}$.

Step 1 Initialize γ, ϖ and the tolerance ε. Set $k = 1$.

Step 2 Use a gradient-based algorithm method (such as Algorithm 5.2.1, Algorithm 5.2.2 and Algorithm 5.2.3 in [9]) to solve Problem $\left(TP_{\gamma,\varpi}\right)_2$ and obtain its optimal solution $\delta_{\mathbf{v}}^{*,k}$ and the corresponding cost $J\left(\delta_{\mathbf{v}}^{*,k}, \mathbf{v}\right)$.

Step 3 If the solution, $x\left(\cdot|\left(\delta_{\mathbf{v}}^{*,k}, \mathbf{v}\right)\right)$, of the (18) corresponding to $\left(\delta_{\mathbf{v}}^{*,k}, \mathbf{v}\right)$ satisfies $h_i\left(x\left(s|\left(\delta_{\mathbf{v}}^{*,k}, \mathbf{v}\right)\right)\right) \geq 0, \forall s \in [0, N], i = 1, \ldots, L$, go to Step 4 and set $k = k+1$. Otherwise, set $\gamma = 10\gamma$ and goto Step 2.

Step 4 If $\left|J\left(\delta_{\mathbf{v}}^{*,k}, \mathbf{v}\right) - J\left(\delta_{\mathbf{v}}^{*,k-1}, \mathbf{v}\right)\right| \leq \varepsilon$, go to Step 5. Otherwise, set $\gamma = 10\gamma$ and $\varpi = \varpi/10$, goto Step 2.

Step 5 $\delta_{\mathbf{v}}^{*,k}$ is an approximate optimal solution of Problem (P) for the fixed $\mathbf{v} \in \boldsymbol{\Pi}$.

Remark 3.1 From Theorem 3.1, we see that the increment of γ as detailed in the loop between Step 2 and Step 3 is a finite process. However, it should be stressed that the validity of Theorem 3.1 is based on the assumption that the optimal solution, $\delta_{\mathbf{v}}^{*,k}$, of Problem $\left(TP_{\gamma,\varpi}\right)$ is a global optimal solution. This does pose a problem, as any gradient based method produces, at best, a local minimizer. Thus, the filled function method proposed in [10] will be used. More specifically, once a local minimizer is obtained, a filled function as described in [10] will be constructed. Then, by minimizing the filled function, its local minimizer will lead to a feasible solution of Problem $\left(TP_{\gamma,\varpi}\right)$ from which a better local minimizer of Problem $\left(TP_{\gamma,\varpi}\right)$ will be obtained. This process is repeated until there exists no local minimizer of the corresponding filled function. For details, see Algorithm 5.1 and Algorithm 5.2 in [10].

In the next section, we will derive an algorithm to solve Problem $\left(TP_{\gamma,\varpi}\right)_1$.

4 Discrete filled function method

Let $\mathbf{e}_{i,j}$ be an element of $\mathbb{R}^{Nm}$ with the i-th component 1 and the remaining components 0, $\tilde{\mathbf{e}}_{i,j}$ be an element of $\mathbb{R}^{Nm}$ with the j-th component -1 and the remaining components 0.

Define

$$\mathbf{D} = \left\{\mathbf{e}_{i,j}, \tilde{\mathbf{e}}_{i,j}, i, j = 1, \ldots, Nm, i \neq j\right\}.$$

Definition 4.1 For any $\mathbf{v} \in \mathbf{\Pi}$, the neighborhood of the integer point $\mathbf{v}$ is defined as $\mathbf{N}(\mathbf{v}) = \{\mathbf{v} + \mathbf{d} : \mathbf{d} \in \mathbf{D}\} \cap \mathbf{\Pi}$.

Definition 4.2 A point $\mathbf{v}^* \in \mathbf{\Pi}$ is said to be a discrete local minimizer of Problem $(TP_{\gamma,\varpi})_1$ if $\bar{J}(\mathbf{v}^*) \leq \bar{J}(\mathbf{v})$ for any $\mathbf{v} \in \mathbf{N}(\mathbf{v}^*) \cap \mathbf{\Pi}$. Furthermore, if $\bar{J}(\mathbf{v}^*) < \bar{J}(\mathbf{v})$ for any $\mathbf{v} \in \mathbf{N}(\mathbf{v}^*) \cap \mathbf{\Pi}$, then $\mathbf{v}^*$ is said to be a strict discrete local minimizer.

Definition 4.3 A point $\mathbf{v}^* \in \mathbf{\Pi}$ is said to be a discrete global minimizer if $\bar{J}(\mathbf{v}^*) \leq \bar{J}(\mathbf{v})$ holds for any $\mathbf{v} \in \mathbf{\Pi}$.

Definition 4.4 A sequence $\left\{\mathbf{v}^i\right\}_{i=1}^{k}$ is called a discrete path in $\mathbf{\Pi}$ between $\mathbf{v}^{1,*} \in \mathbf{\Pi}$ and $\mathbf{v}^{2,*} \in \mathbf{\Pi}$ if the following conditions are satisfied:

1. For any $i = 1, \ldots, k$, $\mathbf{v}^i \in \mathbf{\Pi}$;
2. For any $i \neq j$, $\mathbf{v}^i \neq \mathbf{v}^j$;
3. $\mathbf{v}^1 = \mathbf{v}^{1,*}$, $\mathbf{v}^k = \mathbf{v}^{2,*}$; and
4. $\left\|\mathbf{v}^{i+1} - \mathbf{v}^i\right\| = 2, i = 1, \ldots, k-1$.

We note that $\mathbf{\Pi}$ is a discrete path connected set. That is, for every two different points $\mathbf{v}^1, \mathbf{v}^2$, we can find a path from $\mathbf{v}^1$ to $\mathbf{v}^2$ in $\mathbf{\Pi}$. Clearly, $\mathbf{\Pi}$ is bounded.

Algorithm 4.1 (*Local search*)

1. Choose a $\mathbf{v}_0 \in \mathbf{\Pi}$;
2. If $\mathbf{v}_0$ is a local minimizer, then stop. Otherwise, we search the neighborhood of $\mathbf{v}_0$ and obtain a $\mathbf{v} \in \mathbf{N}(\mathbf{v}_0)$ such that $\bar{J}(\mathbf{v}) < \bar{J}(\mathbf{v}_0)$.
3. Let $\mathbf{v}_0 = \mathbf{v}$, go to Step 2.

After obtaining a local minimizer, we will use a filled function to escape from it. We introduce the following filled function [11] :

$$P(\mathbf{v}, r, \rho) = \frac{1}{r + \bar{J}(\mathbf{v})} \exp\left(-\frac{\left\|\mathbf{v} - \mathbf{v}^{1,*}\right\|^2}{\rho^2}\right). \tag{24}$$

The function $P(\mathbf{v}, r, \rho)$ has the following properties:

Theorem 4.1 *Suppose that* $r + \bar{J}\left(\mathbf{v}^{1,*}\right) > 0$ *and that* $\mathbf{v}^{1,*}$ *is a local minimal solution of* $\bar{J}(\mathbf{v})$. *Then* $\mathbf{v}^{1,*}$ *is a strict local maximal solution of* $P(\mathbf{v}, r, \rho)$ *over* $\mathbf{N}\left(\mathbf{v}^{1,*}\right)$. *That is, for any* $\mathbf{v} \in \mathbf{N}\left(\mathbf{v}^{1,*}\right)$, $P(\mathbf{v}, r, \rho) < P\left(\mathbf{v}^{1,*}, r, \rho\right)$.

Proof Since $\mathbf{v}^{1,*}$ is a local minimal solution of $\bar{J}(\mathbf{v})$, it follows that for any $\mathbf{d} \in \mathbf{D}$, if $\mathbf{v}^{1,*} + \mathbf{d} \in \mathbf{N}\left(\mathbf{v}^{1,*}\right)$, then

$$\bar{J}\left(\mathbf{v}^{1,*} + \mathbf{d}\right) \geq \bar{J}\left(\mathbf{v}^{1,*}\right),$$

and hence

$$r + \bar{J}\left(\mathbf{v}^{1,*} + \mathbf{d}\right) \geq r + \bar{J}\left(\mathbf{v}^{1,*}\right) > 0.$$

Thus,

$$\frac{1}{r + \bar{J}\left(\mathbf{v}^{1,*} + \mathbf{d}\right)} \exp\left(-\frac{\|\mathbf{d}\|^2}{\rho^2}\right) < \frac{1}{r + \bar{J}\left(\mathbf{v}^{1,*}\right)},$$

i.e.,

$$P\left(\mathbf{v}^{1,*} + \mathbf{d}, r, \rho\right) < P\left(\mathbf{v}^{1,*}, r, \rho\right). \qquad \square$$

Let $\bar{J}^{up}$ be an upper bound for the function value of $\bar{J}(\mathbf{v})$ over $\mathbf{\Pi}$. We have the following theorem.

Theorem 4.2 *Suppose that the parameters r, ρ are chosen such that*

$$r + \bar{J}\left(\mathbf{v}^{1,*}\right) > 0, \tag{25}$$

$$\rho^2 \ln \frac{r + \bar{J}^{up}}{r + \bar{J}\left(\mathbf{v}^{1,*}\right)} < 1, \tag{26}$$

where $\bar{J}^{up}$ is an upper bound of $\bar{J}(\mathbf{v})$ in $\mathbf{\Pi}$. Furthermore, for any $\mathbf{v}^1, \mathbf{v}^2 \in \mathbf{\Pi}$, suppose $\bar{J}\left(\mathbf{v}^1\right) \geq \bar{J}\left(\mathbf{v}^{1,*}\right)$, $\bar{J}\left(\mathbf{v}^2\right) \geq \bar{J}\left(\mathbf{v}^{1,*}\right)$ and $\left\|\mathbf{v}^1 - \mathbf{v}^{1,*}\right\|^2 > \left\|\mathbf{v}^2 - \mathbf{v}^{1,*}\right\|^2$. Then,

$$P\left(\mathbf{v}^1, r, \rho\right) < P\left(\mathbf{v}^2, r, \rho\right). \tag{27}$$

Proof Since

$$\frac{r + \bar{J}\left(\mathbf{v}^2\right)}{r + \bar{J}\left(\mathbf{v}^1\right)} \leq \frac{r + \bar{J}^{up}}{r + \bar{J}\left(\mathbf{v}^{1,*}\right)} < \exp\left(1/\rho^2\right) < \exp\left(\frac{\left\|\mathbf{v}^1 - \mathbf{v}^{1,*}\right\|^2 - \left\|\mathbf{v}^2 - \mathbf{v}^{1,*}\right\|^2}{\rho^2}\right),$$

it follows that (27) is satisfied. □

Let the parameters r, ρ be chosen to satisfy the conditions (25), (26). If $\bar{J}(\mathbf{v})$ is not a constant with respect to $\mathbf{v} \in \mathbf{\Pi}$, we choose a sufficiently small h which satisfies

$$0 < h \leq \left|\bar{J}\left(\mathbf{v}^1\right) - \bar{J}\left(\mathbf{v}^2\right)\right|. \tag{28}$$

for any $\mathbf{v}^1, \mathbf{v}^2 \in \mathbf{\Pi}$ such that $\bar{J}\left(\mathbf{v}^1\right) \neq \bar{J}\left(\mathbf{v}^2\right)$.

Theorem 4.3 *Suppose that*

$$0 < r + \bar{J}\left(\mathbf{v}^{1,*}\right) < h. \tag{29}$$

Then, $P\left(\mathbf{v}^1 + d, r, \rho\right) < 0$ if and only if $\bar{J}\left(\mathbf{v}^1 + d\right) < \bar{J}\left(\mathbf{v}^{1,*}\right)$.

Proof Suppose $P\left(\mathbf{v}^1 + d, r, \rho\right) < 0$. Then, it follows from (24) that $r + \bar{J}\left(\mathbf{v}^1 + d\right)$. Now, by (29), $r + \bar{J}\left(\mathbf{v}^{1,*}\right) > 0$. Thus, $\bar{J}\left(\mathbf{v}^1 + d\right) < \bar{J}\left(\mathbf{v}^{1,*}\right)$.

To prove the "only if" statement, we assume that $\bar{J}\left(\mathbf{v}^1 + d\right) < \bar{J}\left(\mathbf{v}^{1,*}\right)$. Then, by (28), we see that $\bar{J}\left(\mathbf{v}^{1,*}\right) - \bar{J}\left(\mathbf{v}^1 + d\right) \geq h$. Thus,

$$r + \bar{J}\left(\mathbf{v}^1 + d\right) \leq r + \bar{J}\left(\mathbf{v}^{1,*}\right) - h < 0.$$

Therefore, $P\left(\mathbf{v}^1 + d, r, \rho\right) < 0$. □

Theorem 4.4 *Let r, ρ satisfy the conditions (29), (26). Then, any discrete local minimal solution of the filled function $P(\mathbf{v}, r, \rho)$ over $\mathbf{\Pi}$ is in the set $\{\mathbf{v} \in \mathbf{\Pi} : P(\mathbf{v}, r, \rho) < 0\}$.*

Proof Suppose the conclusion was false. Then, $P(\mathbf{v}^*, r, \rho) \geq 0$, where $\mathbf{v}^*$ is a minimal solution of $P(\mathbf{v}, r, \rho)$. Thus,

$$\bar{J}\left(\mathbf{v}^*\right) \geq \bar{J}\left(\mathbf{v}^{1,*}\right) \tag{30}$$

by Theorem 4.3. We claim that there exists a $\mathbf{d}^* \in \mathbf{D}$ such that

$$\left\|\mathbf{v}^* + \mathbf{d}^* - \mathbf{v}^{1,*}\right\| > \left\|\mathbf{v}^* - \mathbf{v}^{1,*}\right\|. \tag{31}$$

To establish this claim, we note that $\left\|\mathbf{v}^* - \mathbf{v}^{1,*}\right\|^2 = \sum_{i=1}^{n}\left(\mathbf{v}_i^* - \mathbf{v}_i^{1,*}\right)^2$. There are two cases. (i) there exists some i, j such that $\mathbf{v}_i^* - \mathbf{v}_i^{1,*} > 0$ and $\mathbf{v}_j^* - \mathbf{v}_j^{1,*} < 0$. (ii) $\mathbf{v}_k^* - \mathbf{v}_k^{1,*} \geq 0$ or $\mathbf{v}_k^* - \mathbf{v}_k^{1,*} \leq 0$ for all $1 \leq k \leq n$. For this case, let $i = \max_{1\leq k\leq n}\left|\mathbf{v}_k^* - \mathbf{v}_k^{1,*}\right|$ and $j = \min_{1\leq k\leq n}\left|\mathbf{v}_k^* - \mathbf{v}_k^{1,*}\right|$. Now, by choosing $\mathbf{d}^* = \mathbf{d}_{i,-j}$, we establish the claim. To proceed further, since $\mathbf{v}^*$ is a minimal solution of $P(\mathbf{v}, r, \rho)$, $P(\mathbf{v}^* + \mathbf{d}^*, r, \rho) \geq P(\mathbf{v}^*, r, \rho) \geq 0$. Thus,

$$\bar{J}\left(\mathbf{v}^* + \mathbf{d}^*\right) \geq \bar{J}\left(\mathbf{v}^{1,*}\right) \tag{32}$$

by Theorem 4.3. By virtue of (30), (31) and (32), it follows from Theorem 4.2 that $P(\mathbf{v}^* + \mathbf{d}^*, r, \rho) < P(\mathbf{v}^*, r, \rho)$. This is a contradiction as $\mathbf{v}^*$ is a minimal solution of $P(\mathbf{v}, r, \rho)$. Thus, the conclusion of the theorem must be satisfied. □

From Theorem 4.3, we know that

$$\{\mathbf{v} \in \mathbf{\Pi} : P(\mathbf{v}, r, \rho) < 0\} = \left\{\mathbf{v} \in \mathbf{\Pi} : \bar{J}(\mathbf{v}) < \bar{J}\left(\mathbf{v}^{1,*}\right)\right\}.$$

On this basis, we can construct an algorithm which is based on the following idea. Choose an initial point $\mathbf{v}^1 \in \mathbf{\Pi}$ and use Algorithm 4.1 to find a local minimal $\mathbf{v}^{1,*}$. Then, we construct a filled function to find its local solution. If we can find a point $\mathbf{v}^2$ such that $P\left(\mathbf{v}^2, r, \rho\right) < 0$, then we use $\mathbf{v}^2$ as a new initial point and repeat the above process. Otherwise, we consider $\mathbf{v}^{1,*}$ as a minimizer of $\bar{J}(\mathbf{v})$ over $\mathbf{\Pi}$.

Algorithm 4.2 (*Discrete filled function method*)

1. Take an initial point $\mathbf{v}^1 \in \mathbf{\Pi}$ and let $\Gamma = \left\{\mathbf{v}^1\right\}$.
2. From $\mathbf{v}^1$, use Algorithm 4.1 to find a local minimizer $\mathbf{v}^{1,*}$ of $\bar{J}(\mathbf{v})$ over $\mathbf{\Pi}$. If it has been computed in a local search, we add it to the set Γ.
3. Construct a filled function

$$P(\mathbf{v}, r, \rho) = \frac{1}{r + \bar{J}(\mathbf{v})} \exp\left(-\frac{\left\|\mathbf{v} - \mathbf{v}^{1,*}\right\|^2}{\rho^2}\right)$$

   where r, ρ satisfy the conditions (29) and (26). Use Algorithm 4.1 to find its local minimizer. For each $\mathbf{v} \in \mathbf{\Pi}$, if it has been computed in a local search, we add it to the set Γ. If we find a $\mathbf{v}^2 \in \mathbf{\Pi}$ such that $P\left(\mathbf{v}^2, r, \rho\right) < 0$, let $\mathbf{v}^1 = \mathbf{v}^2$ and go to Step 2. In the local search, if we find a $\mathbf{v}^2 \in \Gamma$, then go back to the father point $\mathbf{v}$ and choose another direction $\mathbf{d} \in \mathbf{D}$ to find a local minimizer. If we cannot find any $\mathbf{v}^2 \in \mathbf{\Pi}$ such that $P\left(\mathbf{v}^2, r, \rho\right) < 0$ or there is no feasible direction to search in, then $\mathbf{v}^{1,*}$ is an optimal solution of $\bar{J}(\mathbf{v})$ over $\mathbf{\Pi}$.

In Algorithm 4.2, we use the set Γ to avoid finding a point which may have been computed repeatedly in the search for a local minimizer of the discrete filled function since any two points in $\mathbf{\Pi}$ are path-connected.

5 Numerical example

In this section, we will apply the algorithms developed in Sects. 3 and 4 to a test problem.

Example 5.1 In this example, we will study the optimal train control problem, which was first presented in [2] and re-considered in [5]. The dynamical system is

$$\begin{aligned}\dot{x}_1 &= x_2\\ \dot{x}_2 &= \varphi(x_2)\,u_1 + \zeta_2 u_2 + \rho(x_2),\end{aligned}$$

where x_1 is the distance along the track, x_2 is the speed of the train, u_1 is the fuel setting and u_2 models the deceleration applied to the train by the brakes. The function

$$\varphi(x_2) = \begin{cases} \zeta_1/x_2, & x_2 \geq \zeta_3 + \zeta_4,\\ \zeta_1/\zeta_3 + \eta_1 (x_2 - (\zeta_3 - \zeta_4))^2 & \\ \quad + \eta_2 (x_2 - (\zeta_3 - \zeta_4))^3, & \zeta_3 - \zeta_4 \leq x_2 \leq \zeta_3 + \zeta_4,\\ \zeta_1/\zeta_3, & x_2 \leq \zeta_3 - \zeta_4,\end{cases}$$

where

$$\eta_1 = \zeta_1 \left\{ \left\{ \frac{1}{\zeta_3 + \zeta_4} - \frac{1}{\zeta_3} \right\} \frac{3}{4\zeta_4^2} + \frac{1}{2\zeta_4 (\zeta_3 + \zeta_4)^2} \right\}$$

and

$$\eta_2 = \zeta_1 \left\{ - \left\{ \frac{1}{\zeta_3 + \zeta_4} - \frac{1}{\zeta_3} \right\} \frac{3}{4\zeta_4^3} - \frac{1}{4\zeta_4^2 (\zeta_3 + \zeta_4)^2} \right\},$$

represents the tractive effort of the locomotive. The function ρ is the resistive acceleration due to friction, given by

$$\rho(x_2) = \zeta_5 + \zeta_6 x_2 + \zeta_7 x_2^2,$$

ζ_i, $i = 1, \ldots, 7$, are constants with given values $\zeta_1 = 1.5$, $\zeta_2 = 1$, $\zeta_3 = 1.4$, $\zeta_4 = 0.1$, $\zeta_5 = -0.015$, $\zeta_6 = -0.00003$ and $\zeta_7 = -0.000006$. The initial state is $\mathbf{x}(0) = (0, 0)^\top$ and the discrete-valued control satisfies $\mathbf{u} = [u_1, u_2] \in \mathbf{U} = \left\{(1, 0)^\top, (0, 0)^\top, (0, -1)^\top\right\}$. Our aim is to find a switching sequence of discrete-valued control, such that $x_1(1500) = 18000$, $x_2(1500) = 0$, and the switching times such that the fuel cost

$$J_0(\mathbf{u}) = \int_0^{1500} u_1 dt$$

is minimized. In this problem, we assume that the maximum number of switchings is 7. Since $x_2(t)$ is the speed, $x_2(t) \geq 0$, for all $t \in [0, 1500]$.

For this optimal control problem, it was solved via a time scaling transform in [5]. In this paper, we will use the discrete filled function method to find the optimal switching sequence and the optimal control software MISER 3.3 to solve Problem $\left(TP_{\gamma,\varpi}\right)_2$.

We choose

$$r = \frac{h}{2} - \bar{J}\left(v^{1,*}\right), \ h = 0.01, \ \bar{J}^{up} = 3000,$$
$$\frac{1}{\rho^2} = 1 + \ln \frac{\bar{J}^{up} - \bar{J}\left(v^{1,*}\right) + h/2}{h/2}.$$

In the local search of the switching sequence of a discrete control, if the integration of a state system or costate system exceeds a given constant, we will assign a large value to the cost corresponding to this sequence. We use MISER as a sub-program to solve this optimal control problem. The cost obtained is 202.4759, which is slightly less than 202.6704 obtained in [5]. The duration of the time used by the control $(0, -1)^\top$ is 2.47197, which is also slightly less than that obtained in [5], which is more than 2.5. All obtained results are depicted in Figs. 1–5. Figure 1 depicts the time scaling transformation of $t\ (s)$ against s. Figures 2 and 3 depict the optimal state $x_1\ (t)$ and $x_2\ (t)$ against the time t. Figures 4 and 5 depict the optimal control $u_1\ (t)$ and $u_2\ (t)$ against the time t.

For the method reported in [5], it is basically a local method. Thus, a good initial guess is needed in the optimization process so as to obtain a good local optimal solution. Also, it cannot be ensured that the assumption on the maximum number of switchings is satisfied. For the method presented in this paper, we can solve the problem from any initial point as a discrete filled function has been incorporated in the algorithm. For our algorithm, once a local minimal solution is obtained, we will search for a local minimal solution of the corresponding filled function. Using this as an initial guess for the next optimization, we will obtain a better local optimal solution. This process is repeated until there exists no local optimal solution of the corresponding discrete filled function. Furthermore, we can be assured that the solution obtained will satisfy the assumption on the maximum number of switchings.

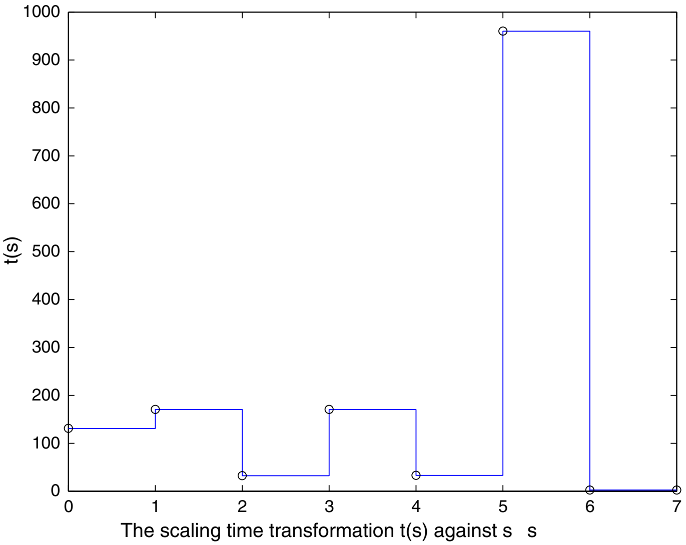

Fig. 1 The profile of the enhancing transformation $t(s)$ versus to s

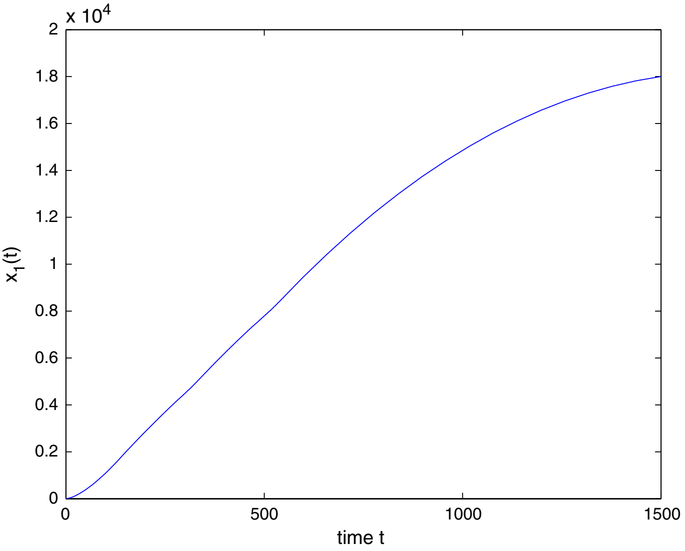

Fig. 2 The profile of optimal state x_1 (t)

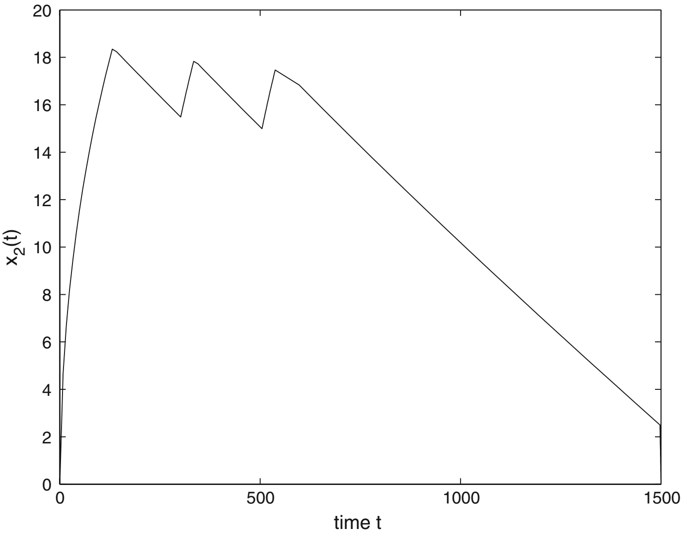

Fig. 3 The profile of optimal state x_2 (t)

6 Conclusion

In this paper, we developed a new computational method to solve a discrete-valued optimal control problem with the maximum number of switchings being prefixed as a two-level optimization problem. In the first level, we use MISER 3.3 to solve it. In the second level, a discrete filled function method is constructed and then used to solve it. To illustrate the method, a numerical example is solved.

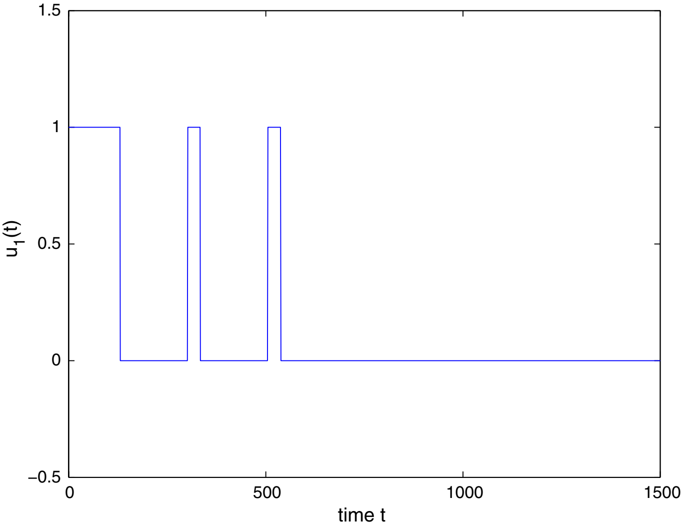

Fig. 4 The profile of optimal control $u_1\ (t)$

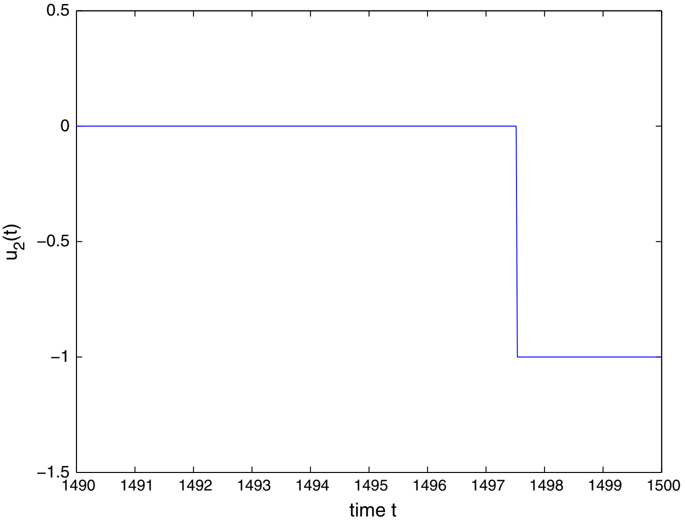

Fig. 5 The profile of optimal control $u_2\ (t)$

Acknowledgements This paper is partially supported by a research grant from the Australian Research Council and a research grant from Chongqing Normal University.

References

1. Gu, Y.H., Wu, Z.Y.: A new filled function method for nonlinear integer programming problem. Appl. Math. Comput. **173**, 938–950 (2006)
2. Howlett, P.: Optimal strategies for the control of a train. Automatica. **32**, 519–532 (1996)

3. Jennings, L.S., Teo, K.L., Fisher, M.E., Goh, C.J.: MISER Version 3, Optimal Control Software, Theory and User Manual. http://www.maths.uwa.edu.au/~les/miser3.3.html. Department of Mathematics, University of Western Australia (2005)
4. Lee, H.W.J., Jennings, L.S., Teo, K.L., Rehbock, V.: Control parametrization enhancing technique for time optimal control problems. Dynam. Syst. Appl. **6**, 243–262 (1997)
5. Lee, H.W.J., Teo, K.L., Rehbock, V., Jennings, L.S.: Control parametrerization enhancing technique for optimal discrete-valued control problems. Automatica. **35**, 1401–1407 (1999)
6. Rehbock, V., Caccetta, L.: Two defence applications involving discrete valued optimal control. ANZIAM J. **44**(E), E33–E54 (2002)
7. Stewart, D.E.: A numerical algorithm for optimal control problems with switching costs. J. Austr. Math. Soc. Ser. B. **34**, 212–228 (1992)
8. Teo, K.L., Jennings, L.S.: Nonlinear optimal control problems with continuous state inequality constraints. J. Optim. Theory Appl. **63**, 1–22 (1989)
9. Teo, K.L., Goh, C.J., Wong, K.H.: A Unified Computational Approach to Optimal Control Problems. Longman Scientific & Technical, Essex, England (1991)
10. Wu, C.Z., Teo, K.L.: Global impulsive optimal control computation. J. Indust. Manage. Optim. **2**, 435–450 (2006)
11. Zhu, W.: An approximate algorithm for nonlinear integer programming. Appl. Math. Comput. **93**, 183–193 (1997)