Delft University of Technology

Delft Center for Systems and Control

Technical report 16-001

Receding-horizon control for max-plus linear systems with discrete actions using optimistic planning*

J. Xu, L. Buşoniu, T. van den Boom, and B. De Schutter

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

J. Xu, L. Buşoniu, T. van den Boom, and B. De Schutter, "Receding-horizon control for max-plus linear systems with discrete actions using optimistic planning," *Proceedings of the 13th International Workshop on Discrete Event Systems*, Xi'an, China, pp. 398–403, May–June 2016.

Delft Center for Systems and Control
Delft University of Technology
Mekelweg 2, 2628 CD Delft
The Netherlands
phone: +31-15-278.24.73 (secretary)
URL: `https://www.dcsc.tudelft.nl`

*This report can also be downloaded via `https://pub.deschutter.info/abs/16_001.html`

Receding-Horizon Control for Max-Plus Linear Systems with Discrete Actions Using Optimistic Planning

Jia Xu*, Lucian Buşoniu†, Ton van den Boom*, Bart De Schutter*
*Delft Center for Systems and Control
Delft University of Technology
Mekelweg 2, 2628 CD Delft, The Netherlands
Email: {j.xu-3, a.j.j.vandenboom, b.deschutter}@tudelft.nl
†Automation Department
Technical University of Cluj-Napoca
Dorobantilor 71-73, 400609 Cluj-Napoca, Romania
Email: lucian@busoniu.net

***Abstract*—This paper addresses the infinite-horizon optimal control problem for max-plus linear systems where the considered objective function is a sum of discounted stage costs over an infinite horizon. The minimization problem of the cost function is equivalently transformed into a maximization problem of a reward function. The resulting optimal control problem is solved based on an optimistic planning algorithm. The control variables are the increments of system inputs and the action space is discretized as a finite set. Given a finite computational budget, a control sequence is returned by the optimistic planning algorithm. The first control action or a subsequence of the returned control sequence is applied to the system and then a receding-horizon scheme is adopted. The proposed optimistic planning approach allows us to limit the computational budget and also yields a characterization of the level of near-optimality of the resulting solution. The effectiveness of the approach is illustrated with a numerical example. The results show that the optimistic planning approach results in a lower tracking error compared with a finite-horizon approach when a subsequence of the returned control sequence is applied.**

I. Introduction

Complex discrete-event systems (DES) such as production systems, railway networks, logistic systems, consist of a finite number of resources (e.g., machines, railway tracks) shared by several users (e.g., workpieces, trains) all of which pursue some common goal (e.g., the assembly of products, transportation of people or goods). The state of such systems evolves in time by the occurrence of asynchronous events (e.g., the start of a processing step, the departure or arrival of a train). In general, DES lead to nonlinear descriptions in conventional algebra. However, there exists a subclass of DES for which we can get a "linear" model in the max-plus algebra [1], [2] whose basic operations are maximization and addition. These systems are called max-plus linear (MPL) systems. Many results have been achieved for modeling and control of MPL systems, see [1]–[7] and the references therein. In particular, finite-horizon control problems for MPL systems are considered in [8]–[10].

In this paper, we consider the optimal control problem for MPL systems with discrete control actions. Sometimes discrete control actions are indeed required in practice. For example, for a manufacturing system it could happen that the raw materials are required to be fed to the manufacturing cell at 6 or 8 hours intervals; or for a railway network the departure times of trains might only be selected as multiples of 15 minutes. These constraints lead to discrete variables. In the given optimal control problem, the objective function is a sum of discounted stage costs over an infinite horizon. Our goal is then to design a control sequence optimizing the infinite-horizon discounted objective function. The approach in this paper is based on optimistic planning algorithms introduced below.

Optimistic planning is a class of planning algorithms originating in artificial intelligence applying the ideas of optimistic optimization [11]. This class of algorithms works for discrete-time systems with general nonlinear (deterministic or stochastic) dynamics and discrete control actions. Based on the current system state, a control sequence is obtained by optimizing an infinite-horizon sum of discounted bounded stage costs (or the expectation of these costs for the stochastic case). Optimistic planning uses a receding-horizon scheme and provides a characterization of the relationship between the computational budget and near-optimality. In [12], three types of optimistic planning algorithms have been reviewed, i.e., optimistic planning for deterministic systems (OPD) [13], open-loop optimistic planning [14], and optimistic planning for sparsely stochastic systems [15]. Moreover, in [12] the theoretical guarantees on the performance of these algorithm are also provided. Recently, optimistic planning has been used for nonlinear networked control systems [16], and nonlinear switched systems [17]. In order to limit computations, optimistic planning with a limited number of action switches has been introduced in [18]. Therefore, optimistic planning can be used for optimal control of very general nonlinear discrete-time systems and in addition it is able to deal with uncertainties

because of the infinite search space and a finite computational budget.

In our previous related paper [19], we use optimistic optimization to solve the finite-horizon optimal control problem for MPL systems with continuous control inputs. In this paper, we propose to apply optimistic planning to solve the infinite-horizon optimal control problem for MPL systems where the action space is discretized as a finite set. Note that although the evolution of MPL systems is event-driven in contrast to time-driven as in a discrete-time system, optimistic planning can still be applied because of the analogy between descriptions of MPL systems and conventional linear time-invariant discrete-time systems. Also note that considering an infinite-horizon discounted objective function is more flexible than selecting a fixed finite-horizon objective function since the prediction horizon does not have to be fixed a priori. The length of the returned control sequence varies depending on the computational budget, the complexity of the problem, and the discount factor. Based on the standard geometric series, discounting is a simple way to obtain finite values for the total sum of stage costs over an infinite horizon. This is very convenient for comparing different infinite-length control sequences.

This paper is organized as follows. In Section II, some preliminaries regarding max-plus linear systems and optimistic planning are given. In Section III, the formulation of the infinite-horizon discounted optimal control problem for max-plus linear systems and the optimistic planning based approach are presented. Next an example is included in Section IV to illustrate the performance of the proposed approach. Finally, Section V concludes the paper.

II. Preliminaries and background

A. Max-plus linear systems

Define $\varepsilon = -\infty$ and $\mathbb{R}_\varepsilon = \mathbb{R} \cup \{\varepsilon\}$. The max-plus-algebraic addition ($\oplus$) and multiplication ($\otimes$) are defined as [1]:

$$x \oplus y = \max(x, y), \quad x \otimes y = x + y$$

for any $x, y \in \mathbb{R}_\varepsilon$. For matrices $A, B \in \mathbb{R}_\varepsilon^{m\times n}$ and $C \in \mathbb{R}_\varepsilon^{n\times l}$, we define

$$\big[A \oplus B\big]_{ij} = a_{ij} \oplus b_{ij} = \max(a_{ij}, b_{ij})$$
$$\big[A \otimes C\big]_{ij} = \bigoplus_{k=1}^{n} a_{ik} \otimes c_{kj} = \max_{k=1,\ldots,n}(a_{ik} + c_{kj})$$

for all i, j. The zero matrix $\mathcal{E}$ in max-plus algebra has all its entries equal to ε. The identity matrix E in max-plus algebra has the diagonal entries equal to 0 and the other entries equal to ε.

Consider a max-plus linear (MPL) system [20]

$$x(k+1) = A \otimes x(k) \oplus B \otimes u(k) \tag{1}$$
$$y(k) = C \otimes x(k) \tag{2}$$

with the system matrices $A \in \mathbb{R}_\varepsilon^{n_x\times n_x}, B \in \mathbb{R}_\varepsilon^{n_x\times n_u}, C \in \mathbb{R}_\varepsilon^{n_y\times n_x}$, where n_x is the number of states, n_u is the number of inputs, and n_y is the number of outputs. The index $k \in \{0, 1, \ldots\}$ is called the event counter. The components of $u(k)$, $x(k)$, and $y(k)$ are typically input, state, output event occurrence times. For example, if the MPL system is a model of a manufacturing system, $u(k), x(k)$ and $y(k)$ are the k-th feeding times of raw materials, the k-th starting times of the production processes, and the k-th completion times for the end products. Note that the event times can easily be measured; so we consider the case of full state information. Since the inputs represent event times, a typical constraint is that the control sequence should be nondecreasing, i.e.,

$$u(k+1) - u(k) \geq 0 \quad \forall\ k \geq 0. \tag{3}$$

B. Optimistic planning for deterministic systems

Optimistic planning for deterministic systems (OPD) [11], [13] is an algorithm that solves an optimal control problem for discrete-time deterministic systems described by an equation of the form

$$x_{k+1} = f(x_k, u_k)$$

with discrete control inputs $u_k \in U \triangleq \{u^1, \ldots, u^M\}$. In this section, k is a time counter[1]. Given the initial state x_0, OPD designs a control sequence $\boldsymbol{u} = (u_0, u_1, \ldots)$ maximizing the following infinite-horizon discounted reward function:

$$\bar{J}(\boldsymbol{u}, x_0) = \sum_{k=0}^{\infty} \gamma^k r_{k+1} \tag{4}$$

where $r_{k+1} \in [0, 1]$ is the reward for the transition from x_k to x_{k+1} as a result of u_k and where $\gamma \in (0, 1)$ is the discount factor that is often used in the fields of dynamic programming and reinforcement learning and expresses the difference in importance between future costs and present costs. The value of γ is usually selected close to 1. The optimal value of (4) is denoted as $\bar{J}^*(x_0) = \max_{\boldsymbol{u}} \bar{J}(\boldsymbol{u}, x_0)$.

For a given initial state, OPD explores the space of all possible control sequences $\boldsymbol{u}$. Define $\boldsymbol{u}_d = (u_0, \ldots, u_{d-1})$ as a length d sequence with $d \in \{1, 2, \ldots\}$ and define $\boldsymbol{u}|_d$ as any infinite-length sequence of which the first d components coincide with $\boldsymbol{u}_d$. For any x_0, each $\boldsymbol{u}_d$ determines a state sequence $x_1, \ldots, x_d$. Define

$$v(\boldsymbol{u}_d) = \sum_{k=0}^{d-1} \gamma^k r_{k+1} \tag{5}$$
$$b(\boldsymbol{u}_d) = v(\boldsymbol{u}_d) + \frac{\gamma^d}{1-\gamma}. \tag{6}$$

The value $v(\boldsymbol{u}_d)$ is the sum of discounted rewards along the trajectory starting from the initial state x_0 and applying the control sequence $\boldsymbol{u}_d$, and provides a lower bound of the value $\bar{J}(\boldsymbol{u}|_d, x_0)$ for any $\boldsymbol{u}|_d$. On the other hand, note that $r_k \in [0, 1]$;

[1]In order to distinguish between the event counter and time counter, we use the notation $x(k)$ when k is an event counter and x_k when k is a time counter.

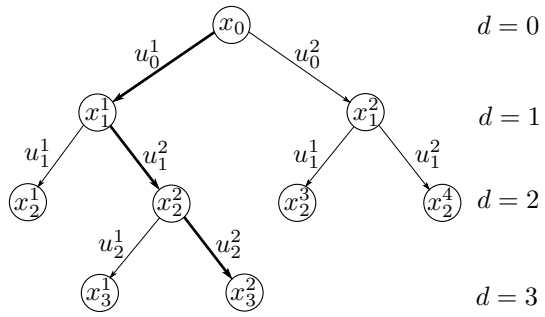

Fig. 1. The tree representation of OPD with $M = 2$, i.e., $U = \{u^1, u^2\}$. The root node at depth $d = 0$ denotes the initial state x_0. Each edge starting from a node at depth d corresponds to a control action u_d^i, $i = 1, \ldots, M$. Each node corresponds to a reachable state x_d^i, $i = 1, \ldots, M^d$. The depth d corresponds to the time step. Any node at depth d is reached by a unique sequence $\boldsymbol{u}_d$ (e.g., the thick line for node x_3^2) starting from x_0.

hence,

$$\begin{aligned}\bar{J}(\boldsymbol{u}|_d, x_0) &= v(\boldsymbol{u}_d) + \sum_{k=d}^{\infty} \gamma^k r_{k+1} \\ &\leq v(\boldsymbol{u}_d) + \sum_{k=d}^{\infty} \gamma^k \cdot 1 \\ &\leq v(\boldsymbol{u}_d) + \frac{\gamma^d}{1-\gamma}.\end{aligned}$$

So $b(\boldsymbol{u}_d)$ provides an upper bound of $\bar{J}(\boldsymbol{u}|_d, x_0)$ for any $\boldsymbol{u}|_d$.

The search process of OPD over the space of all possible control sequences $\boldsymbol{u}$ can be represented as a tree exploration process, as illustrated in Fig. 1. Nodes of the tree correspond to reachable states; in particular, the root node is the initial state x_0. Edges of the tree correspond to the possible control actions. Each node at some depth d is reached by a unique path through the tree, i.e., each node corresponds to a unique control sequence $\boldsymbol{u}_d = (u_0, \ldots, u_{d-1})$. Expanding a node means adding its M children to the current tree, i.e., generating transitions and rewards as well as computing the v and b-values for the M children. Given a finite number of node expansions, at each step, OPD always expands the most promising leaf[2], i.e., the control sequence $\boldsymbol{u}_d$ with the largest upper bound $b(\boldsymbol{u}_d)$. The algorithm terminates if the given number of node expansions n has been reached. Finally, the algorithm returns the control sequence $\boldsymbol{u}_{d'}^* = (u_0^*, u_1^*, \ldots, u_{d'-1}^*)$ that maximizes the lower bound v where d' is the length of the returned optimal control sequence. The process of OPD is summarized in Algorithm 1.

Define the set of near-optimal nodes at depth d as follows:

$$\mathcal{T}_d^* = \left\{ \boldsymbol{u}_d \middle| \bar{J}^*(x_0) - v(\boldsymbol{u}_d) \leq \frac{\gamma^d}{1-\gamma} \right\}.$$

OPD only expands the nodes in $\mathcal{T}_d^*$, $d = 0, 1, 2, \ldots$, so the number of nodes in $\mathcal{T}_d^*$, denoted as $|\mathcal{T}_d^*|$, determines the efficiency of the algorithm. Define the asymptotic branching factor $\kappa \in [1, M]$ as $\kappa = \limsup_{d\to\infty} |\mathcal{T}_d^*|^{1/d}$, which characterizes the complexity of the problem. The following theorem summarizes the near-optimality analysis presented in [11], [13], [16].

[2]A leaf of a tree is a node with no children.

Algorithm 1 Optimistic planning for deterministic systems

Input: initial state x_0, action space $U = \{u^1, \ldots, u^M\}$, number of node expansions n
Initialize: $\mathcal{T} \leftarrow \{x_0\}$
expand the root node by adding its M children to $\mathcal{T}$
$t \leftarrow 1$
while $t < n$
  expand the leaf with largest b-value
  $t \leftarrow t + 1$
end while
return $\boldsymbol{u}_{d'}^* = \arg\max_{\boldsymbol{u}_d \in \mathcal{L}(\mathcal{T})} v(\boldsymbol{u}_d)$ where $\mathcal{L}(\mathcal{T})$ is the set of leaves of $\mathcal{T}$

Theorem 1: Let the initial state x_0 and the number of node expansions n be given.

(i) Let $\boldsymbol{u}_{d'}^*$ be the sequence returned by the OPD algorithm and let $\boldsymbol{u}^*|_{d'}$ be any infinite-length sequence of which the first d' components coincide with $\boldsymbol{u}_{d'}^*$. Then we have $\bar{J}^*(x_0) - \bar{J}(\boldsymbol{u}^*|_{d'}, x_0) \leq b(\boldsymbol{u}_{d'}^*) - v(\boldsymbol{u}_{d'}^*) \leq \frac{\gamma^{d'}}{1-\gamma}$.

(ii) If $\kappa > 1$, then $\bar{J}^*(x_0) - \bar{J}(\boldsymbol{u}^*|_{d'}, x_0) = O\Big(n^{-\frac{\log 1/\gamma}{\log \kappa}}\Big)$.

(iii) If $\kappa = 1$, then $\bar{J}^*(x_0) - \bar{J}(\boldsymbol{u}^*|_{d'}, x_0) = O\big(\gamma^{cn}\big)$ where c is a constant. □

Remark 2: Theorem 1(i) provides an a posteriori bound on the near-optimality of the returned control sequence; while Theorem 1(ii)-(iii) imply a prior bounds based on the complexity of the problem. The branching factor κ characterizes the number of nodes that will be expanded by the OPD algorithm. If $\kappa > 1$, then OPD needs the number of expansions $n = O(\kappa^d)$ to reach the depth d in the optimistic planning tree; if $\kappa = 1$, then $n = O(d)$ is required. Thus, $\kappa = 1$ is the ideal case where the number of near-optimal nodes at every depth is bounded by a constant independent of d and the a prior bound on the near-optimality decreases exponentially with n.

OPD uses a receding-horizon scheme, so once $\boldsymbol{u}_{d'}^*$ has been computed, subsequently, only the first component u_0^* of $\boldsymbol{u}_{d'}^*$ is applied to the system, resulting in the state x_1^*. At the next time step, x_1^* is used as the initial state and the whole process is repeated.

III. Optimistic planning for max-plus linear systems

A. Problem statement

In this paper, we consider the optimal control problem for the MPL system (1)-(2). The input $u(k)$ is rewritten as

$$u(k) = u(k-1) + \Delta u(k). \tag{7}$$

We consider the single input case (i.e., $n_u = 1$) for the sake of simplicity; however, an extension to multiple inputs can be made. We assume that the increments $\Delta u(k)$ of the input take values from a given finite set $U \triangleq \{u^1, \ldots, u^M\}$ with M the number of actions and with $u^i \geq 0$ for all i, and where U is called the action space.

Given a reference signal $\{y^{\mathrm{ref}}(k)\}_{k=0}^{\infty}$ with $y^{\mathrm{ref}}(k) \in \mathbb{R}^l$, a typical objective in optimal control for MPL systems is

minimizing the tracking error (e.g., the tardiness $\max(y_j(k) - y_j^{\mathrm{ref}}(k), 0)$) between the output event times and the reference signal, which represents a due date signal. So we consider the following stage cost:

$$\rho(k) = \sum_{j=1}^{n_y} \min\big(\max(y_j(k) - y_j^{\mathrm{ref}}(k), 0), g\big) + \lambda F(\Delta u(k)) \tag{8}$$

where the positive scalar g is introduced to make $\rho(k)$ bounded, and $\lambda > 0$ is a trade-off between the delay of completion times with respect to the due date signal and the feeding rate. For each element u^i of the finite set U, we assign a cost F according to some criterion. If we consider a just-in-time setting, then the smaller the value of $\Delta u(k)$, the larger the value of its cost, i.e., F should be a positive monotonically nonincreasing function of $\Delta u(k)$. For example, assume that $U = \{u^1, u^2\}$, i.e., the next feeding time is after u^1 or u^2 time units and assume that $u^1 < u^2$, then we could have

$$F(\Delta u(k)) = \alpha_i g \quad \text{if } \Delta u(k) = u^i$$

with $\alpha_1 > \alpha_2$ and $\alpha_1 + \alpha_2 = 1$. Another example could be:

$$F(\Delta u(k)) = g - \Delta u(k) \quad \text{with } g \geq \max(U).$$

It is easy to verify that $\rho(k)$ always belongs to the interval $[0, g + \lambda g]$.

Given initial conditions $x(0)$ and $u(-1)$, define an infinite-length control sequence $\Delta \boldsymbol{u} = (\Delta u(0), \Delta u(1), \ldots)$ and the corresponding infinite-horizon discounted cost function of this sequence:

$$J(\Delta \boldsymbol{u}, x(0), u(-1)) = \sum_{k=0}^{\infty} \gamma^k \rho(k+1)$$

Note that we have $J(\Delta \boldsymbol{u}, x(0), u(-1)) \in [0, \frac{g+\lambda g}{1-\gamma}]$.

The infinite-horizon discounted optimal control problem for MPL systems with discrete actions is now defined as follows:

$$\min_{\Delta \boldsymbol{u}} \; J(\Delta \boldsymbol{u}, x(0), u(-1)) = \sum_{k=0}^{\infty} \gamma^k \rho(k+1) \tag{9}$$

subject to (1), (2), (7) and

$$\Delta u(k) \in U \triangleq \{u^1, \ldots, u^M\}, \quad k = 0, 1, \ldots \tag{10}$$

Note that (3) is automatically satisfied since $u^i \geq 0$ for all i.

B. Approach

In order to apply OPD to solve the infinite-horizon discounted optimal control problem (9)-(10), we first define lower and upper bound functions similar to (5) and (6). The bounded stage cost function (8) corresponds to a bounded reward function:

$$r(k) = 1 - \frac{\rho(k)}{g + \lambda g}. \tag{11}$$

Furthermore, $r(k) \in [0, 1]$. The minimization problem (9) can now be translated into the following maximization problem:

$$\max_{\Delta \boldsymbol{u}} \; \bar{J}(\Delta \boldsymbol{u}, x(0), u(-1)) = \sum_{k=0}^{\infty} \gamma^k r(k+1) \tag{12}$$

subject to (1), (2), (7), (8), (10), and (11). (13)

Define

$$\Delta \boldsymbol{u}_d = (\Delta u(0), \ldots, \Delta u(d-1))$$

$$v(\Delta \boldsymbol{u}_d) = \sum_{k=0}^{d-1} \gamma^k r(k+1)$$

$$b(\Delta \boldsymbol{u}_d) = v(\Delta \boldsymbol{u}_d) + \frac{\gamma^d}{1-\gamma}.$$

So $v(\Delta \boldsymbol{u}_d)$ and $b(\Delta \boldsymbol{u}_d)$ provide lower and upper bounds of $\bar{J}(\Delta \boldsymbol{u}|_d, x(0), u(-1))$ for any infinite-length sequence $\Delta \boldsymbol{u}|_d$ of which the first d components coincide with $\Delta \boldsymbol{u}_d$. When applying OPD to solve the problem (12)-(13), the upper bound function b is used to select the most promising control sequence (corresponding to the largest b-value among all leaves of the current tree) to expand. The lower bound function v is used for determining the best control sequence at the end of the algorithm.

Given initial conditions $x(0)$ and $u(-1)$, a reference signal $\{y^{\mathrm{ref}}(k)\}_{k=0}^{\infty}$, and the number of node expansions n, OPD returns a control sequence $\Delta \boldsymbol{u}^*_{d'}$ that maximizes the lower bound v function. The first action of $\Delta \boldsymbol{u}^*_{d'}$ is applied to the system and the whole process is repeated at each event step. In this way, a receding-horizon controller is obtained. The length d' of the returned sequence is the maximum depth reached by the algorithm for the given finite n. According to Theorem 1(i), we have the following corollary for the near-optimality guarantee of the returned control sequence:

Corollary 3: Let

$$\bar{J}^*(x(0), u(-1)) \triangleq \max_{\Delta \boldsymbol{u}} \bar{J}(\Delta \boldsymbol{u}, x(0), u(-1))$$

be the optimal value of the objective function in (12). Let $\Delta \boldsymbol{u}^*|_{d'}$ be any infinite-length sequence of which the first d' components coincide with $\Delta \boldsymbol{u}^*_{d'}$ returned by OPD. Then we have

$$\begin{aligned} &\bar{J}^*(x(0), u(-1)) - \bar{J}(\Delta \boldsymbol{u}^*|_{d'}, x(0), u(-1)) \\ &\quad \leq b(\Delta \boldsymbol{u}^*_{d'}) - v(\Delta \boldsymbol{u}^*_{d'}) \\ &\quad \leq \frac{\gamma^{d'}}{1-\gamma}. \end{aligned}$$

□

OPD applies just the first component of $\Delta \boldsymbol{u}^*_{d'}$ to the system and generates a new control sequence at the next event step. Rather than recomputing a new control sequence at every event step, one can alternatively apply the first subsequence of length $\bar{d}$ of $\Delta \boldsymbol{u}^*_{d'}$ (with $\bar{d} \leq d'$) to the system and recompute the control sequence only every $\bar{d}$ event steps. Namely, once a length $\bar{d}$ control sequence is applied, the next sequence is computed from the predicted state at the

end of the current sequence. Applying sequences of control actions in parallel with running OPD to find the next control sequence is investigated in [21] where conditions under which the algorithm is guaranteed to be feasible in real-time are provided. Recall that d' is the maximum depth reached by the algorithm for the fixed n. In order to obtain a control sequence with a sufficient length, the number of node expansions n should be large enough such that the length of the returned sequence $\Delta \boldsymbol{u}^*_{d'}$ is at least $\bar{d}$. In the worst case, the algorithm will explore all branches of the tree, so n should be larger than $\sum_{k=0}^{\bar{d}-1} M^k + 1$ to generate that at least one path has length $\bar{d}$. However, in general a smaller n can be selected because OPD explores the tree in an efficient way rather than evaluating all actions in the action space at each step of node expansion. We can also add the depth $\bar{d}$ as a new termination rule in OPD. Applying a subsequence of length $\bar{d}$ means that the controller has more time to compute a new control sequence, so we can then increase n. This in general may have positive effect on performance.

C. Relation to Model Predictive Control

From the viewpoint of the receding-horizon scheme, optimistic planning can be seen as a variant of model predictive control (MPC). In MPC, a receding-horizon controller is obtained by repeatedly solving a finite-horizon open-loop optimal control problem and applying the first control input to the system. Using the current system state as the initial state, a control sequence is computed by optimizing an objective function over a finite horizon (prediction horizon). The whole procedure is repeated at the next step when new state measurements are available. Different from MPC, rather than a fixed horizon setting optimistic planning optimizes an infinite-horizon discounted objective function. The length of the returned control sequence is influenced by the computational budget, the value of the discount factor γ, and the complexity of the problem.

IV. Example

Consider the following MPL system from [22]

$$x(k+1) = \begin{bmatrix} \varepsilon & 0 & \varepsilon & 9 \\ 4 & 3 & 4 & 5 \\ 8 & \varepsilon & 2 & 8 \\ 0 & 1 & \varepsilon & \varepsilon \end{bmatrix} \otimes x(k) \oplus \begin{bmatrix} 0 \\ 5 \\ 2 \\ 8 \end{bmatrix} \otimes u(k) \quad (14)$$

$$y(k) = \begin{bmatrix} 6 & 5 & 8 & \varepsilon \end{bmatrix} \otimes x(k). \quad (15)$$

Given a due date signal $y^{\mathrm{ref}}(k) = 50 + 6.5k$, and the initial conditions $x(0) = \begin{bmatrix} 6 & 12 & 9 & 14 \end{bmatrix}^T$ and $u(-1) = 6$, we consider the following stage cost function

$$\rho(k) = \min\big(\max(y(k) - r(k), 0), g\big) + \lambda(g - \Delta u(k)) \quad (16)$$

with $g = 500$, $\lambda = 0.001$, $\Delta u(k) \in U = \{6, 8\}$, $M = 2$.

The optimistic planning based approach is implemented to obtain a receding-horizon controller for the MPL system (14)-(15). In addition, a finite-horizon approach is also implemented for comparison. More specifically, given a fixed finite horizon d_{N}, a full tree[3] is explored from the root node to the depth d_{N}. The finite-horizon approach returns a control sequence that maximizes the following function

$$\bar{J}_{\mathrm{N}} = \sum_{k=0}^{d_{\mathrm{N}}-1} \gamma^k r(k+1)$$

where $\gamma = 0.95$ and r is the reward corresponding to (16).

The difference $y^{\mathrm{ref}} - y$ is used for comparing the optimistic planning approach and the finite-horizon approach. For each approach, we consider both applying the first action only and applying a subsequence of length $\bar{d}$ to the system once an optimal control sequence is obtained. Fig. 2 shows the results of applying the first action only with $n = 100$ for the optimistic planning approach and with $d_{\mathrm{N}} = 10$ for the finite-horizon approach. We can see that the two approaches result in the same tracking error. Fig. 3 shows the results of applying a subsequence of length $\bar{d} = 9$ with $n = 500$ and $d_{\mathrm{N}} = 10$. We can see that in this case the optimistic planning approach gives a lower tracking error than the finite-horizon approach. In addition, for both approaches, the range of tracking errors by applying a subsequence is smaller than that by applying the first action only. Thus, for the considered MPL system (14)-(15), applying a subsequence of length $\bar{d} = 9$ yields better tracking than applying the first action only for both approaches. However, this does not mean that applying a subsequence performs better for any experimental instance.

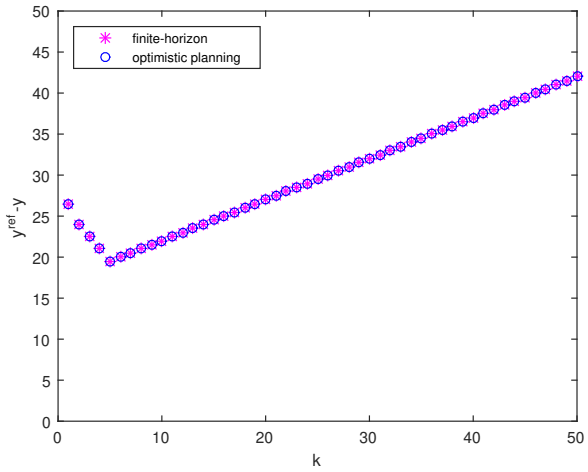

Fig. 2. Tracking error for the closed-loop controlled system when applying the first action only of the returned sequences

V. Conclusions

In this paper, we have considered the infinite-horizon optimal control problem for max-plus linear (MPL) systems. The considered infinite-horizon discounted objective function aims at reducing the tracking error between the output and a reference signal. We have adapted optimistic planning to solve

[3]Here a full tree is a tree in which every node other than the leaves has M children.

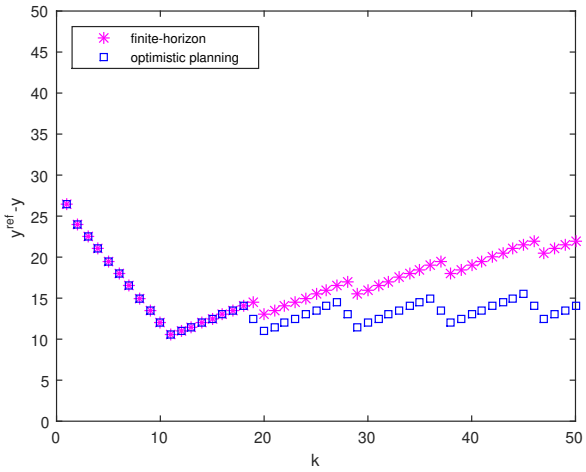

Fig. 3. Tracking error for the closed-loop controlled system when applying the first subsequence of length $\bar{d} = 9$ of the returned sequences

the resulting problem by taking the increments of the inputs as control variables and a discrete action space. Within a limited computational budget, the optimistic planning algorithm returns a control sequence of which the near-optimality can be characterized. In particular a bound can be derived for the difference between the optimal value of the objective function and the near-optimal value corresponding to the returned control sequence. A numerical example has been implemented to assess the effectiveness of the proposed approach. The results show that for the given MPL system the proposed approach yields better tracking than a finite-horizon approach when applying a subsequence of the returned control sequence.

In the future, we will focus on solving the robust optimal control problem for MPL systems with disturbances using (variants of) optimistic planning. We will also explore the infinite-horizon optimal control problem for other discrete-event and hybrid systems such as max-min-plus-scaling and piecewise affine systems.

ACKNOWLEDGMENT

Research supported by the Chinese Scholarship Council and a grant of the Romanian National Authority for Scientific Research, CNCS-UEFISCDI, project number PNII-RU-TE-2012-3-0040.

REFERENCES

[1] F. L. Baccelli, G. Cohen, G. J. Olsder, and J.-P. Quadrat, *Synchronization and Linearity: An Algebra for Discrete Event Systems*. New York: John Wiley & Sons, 1992.

[2] B. Heidergott, G. J. Olsder, and J. W. van der Woude, *Max Plus at Work: Modeling and Analysis of Synchronized Systems*. Princeton, New Jersey: Princeton University Press, 2006.

[3] T. van den Boom and B. De Schutter, "Model predictive control of manufacturing systems with max-plus algebra," in *Formal Methods in Manufacturing*, ser. Industrial Information Technology, J. Campos, C. Seatzu, and X. Xie, Eds. CRC Press, Feb. 2014, ch. 12, pp. 343–380.

[4] E. Menguy, J.-L. Boimond, L. Hardouin, and J.-L. Ferrier, "Just-in-time control of timed event graphs: update of reference input, presence of uncontrollable input," *IEEE Transactions on Automatic Control*, vol. 45, no. 9, pp. 2155–2159, 2000.

[5] B. Cottenceau, L. Hardouin, J. L. Boimond, and J. L. Ferrier, "Model reference control for timed event graphs in dioids," *Automatica*, vol. 37, no. 9, pp. 1451–1458, 2001.

[6] C. A. Maia, C. R. Andrade, and L. Hardouin, "On the control of max-plus linear system subject to state restriction," *Automatica*, vol. 47, no. 5, pp. 988–992, 2011.

[7] L. Houssin, S. Lahaye, and J.-L. Boimond, "Control of (*max*,+)-linear systems minimizing delays," *Discrete Event Dynamic Systems*, vol. 23, no. 3, pp. 261–276, 2013.

[8] I. Necoara, E. C. Kerrigan, B. De Schutter, and T. J. J. van den Boom, "Finite-horizon min-max control of max-plus-linear systems," *IEEE Transactions on Automatic Control*, vol. 52, no. 6, pp. 1088–1093, 2007.

[9] I. Necoara, T. van den Boom, B. De Schutter, and H. Hellendoorn, "Stabilization of max-plus-linear systems using model predictive control: The unconstrained case," *Automatica*, vol. 44, no. 4, pp. 971–981, Apr. 2008.

[10] J. Haddad, B. De Schutter, D. Mahalel, I. Ioslovich, and P.-O. Gutman, "Optimal steady-state control for isolated traffic intersections," *IEEE Transactions on Automatic Control*, vol. 55, no. 11, pp. 2612–2617, 2010.

[11] R. Munos, "From bandits to Monte-Carlo tree search: The optimistic principle applied to optimization and planning," *Foundations and Trends in Machine Learning*, vol. 7, no. 1, pp. 1–130, 2014.

[12] L. Buşoniu, R. Munos, and R. Babuska, "A survey of optimistic planning in Markov decision processes," in *Reinforcement Learning and Adaptive Dynamic Programming for Feedback Control*. Wiley, 2012, pp. 494–516.

[13] J. F. Hren and R. Munos, "Optimistic planning of deterministic systems," in *Proceedings 8th European Workshop on Reinforcement learning*, Villeneuve d'Ascq, France, 2008, pp. 151–164.

[14] S. Bubeck and R. Munos, "Open loop optimistic planning," in *Proceedings 23rd Annual Conference on Learning Theory*, Haifa, Israel, 2010, pp. 27–29.

[15] L. Buşoniu, R. Munos, B. De Schutter, and R. Babuška, "Optimistic planning for sparsely stochastic systems," in *Proceedings 2011 IEEE International Symposium on Adaptive Dynamic Programming and Reinforcement Learning*, Paris, France, 2011, pp. 48–55.

[16] L. Buşoniu, R. Postoyan, and J. Daafouz, "Near-optimal strategies for nonlinear networked control systems using optimistic planning," in *Proceedings American Control Conference*, Washington DC, USA, Jun. 2013, pp. 3020–3025.

[17] L. Buşoniu, M.-C. Bragagnolo, J. Daafouz, and C. Morarescu, "Planning methods for the optimal control and performance certification of general nonlinear switched systems," in *Proceedings 54th IEEE Conference on Decision and Control*, Osaka, Japan, 2015, pp. 3604–3609.

[18] K. Máthé, L. Buşoniu, R. Munos, and B. De Schutter, "Optimistic planning with a limited number of action switches for near-optimal nonlinear control," in *Proceedings 53rd IEEE Conference on Decision and Control*, Los Angeles, California, USA, Dec. 2014, pp. 3518–3523.

[19] J. Xu, B. De Schutter, and T. van den Boom, "Model predictive control for max-plus-linear systems via optimistic optimization," in *Proceedings 12th International Workshop on Discrete Event Systems*, Cachan, France, May 2014, pp. 111–116.

[20] B. De Schutter and T. van den Boom, "Model predictive control for max-plus-linear discrete event systems," *Automatica*, vol. 37, no. 7, pp. 1049–1056, Jul. 2001.

[21] T. Wensveen, L. Buşoniu, and R. Babuška, "Real-time optimistic planning with action sequences," in *Proceedings 20th International Conference on Control Systems and Computer Science*, Bucharest, Romania, 2015, pp. 923–930.

[22] I. Necoara, "Model predictive control for piecewise affine and max-plus-linear systems," Ph.D. dissertation, Delft University of Technology, 2006.