Delft University of Technology

Delft Center for Systems and Control

Technical report 11-011

An approximation approach for identification of stochastic max-plus linear systems*

S.S. Farahani, T. van den Boom, and B. De Schutter

* This report can also be downloaded via https://pub.bartdeschutter.org/abs/11_011

An Approximation Approach for Identification of Stochastic Max-Plus Linear Systems

Samira S. Farahani [*] **Ton van den Boom** [*] **Bart De Schutter** [*]

[*] *Delft Center for Systems and Control, Delft University of Technology, Delft,*
The Netherlands (e-mail: {s.safaeifarahani,t.j.j.vandenboom,b.deschutter}@tudelft.nl)

Abstract: Max-plus linear systems belong to a special class of discrete-event systems that consists of systems with synchronization but no choice. Our focus in this paper is on stochastic max-plus linear systems, i.e., perturbed systems that are linear in the max-plus algebra. One interesting topic is the identification of such systems. Previous works report on a method for identifying the parameters of a state space model for a stochastic max-plus linear system from measured data. However, due to the structure of such systems, this method results in a complex identification problem. Therefore, the aim of this paper is to decrease the computational complexity and the computation time of this problem. To this end, we use an approximation approach that is based on the higher-order raw moments of a random variable. This method results in a less complex problem that can be solved efficiently using gradient search techniques.

Keywords: Stochastic discrete event systems, stochastic max-plus linear systems, identification, approximation, raw moments

1. INTRODUCTION

Discrete-event systems are dynamical event-driven systems in which the state of the system changes due to the occurrence of events, in contrast to continuous variable systems where the progression of time specifies the system state. Discrete-event systems arise in the context of telecommunication networks, railway networks, manufacturing systems, parallel computing, traffic control systems, etc., for which there exist different modeling frameworks such as queuing theory, (extended) state machines, formal languages, automata, temporal logic models, generalized semi-Markov processes, Petri nets, and computer simulation models (Cassandras and Lafortune, 1999; Ho, 1992; Peterson, 1981). One of the important classes of discrete-event systems is the class of the max-plus linear (MPL) systems, which contains discrete-event systems with synchronization but no choice. This leads to a description that is linear in max-plus algebra (Baccelli et al., 1992; Cuninghame-Green, 1979; Heidergott et al., 2006), and that applies to both deterministic and stochastic discrete-event systems. In stochastic systems, processing times and/or transportation times are assumed to be stochastic quantities. More results on stochastic MPL systems such as system analysis, control system, etc., can be found in (Baccelli et al., 1992; Heidergott et al., 2006; Mairesse, 1994; Olsder et al., 1990; Resing et al., 1990).

To use a model for control purposes, the model parameters have to be determined. We concentrate on state space identification for stochastic MPL discrete-event systems. In general, compared to transfer functions, State space models have certain advantages: they explicitly take the initial state of the system into account, they can reveal "hidden" behavior such as unobservable, unstable modes, the extension from SISO to MIMO is more intuitive and elegant for state space models, and the analysis is often easier. Some examples of state space identification methods for *deterministic* MPL systems are present in (De

Schutter et al., 2002; Schullerus and Krebs, 2001a,b; Schullerus et al., 2003).

In this paper, we consider the identification of the *stochastic* MPL systems. In such systems modeling errors, noise, and/or disturbances are present, which results in the perturbation of the system parameters and consequently, the system identification has to take the stochastic properties into account. This problem has been addressed in (van den Boom et al., 2003) where under quite general conditions the resulting identification problem can be solved using the gradient search techniques. However, when the order of the stochastic system increases, the computational complexity increases drastically and hence the method proposed in (van den Boom et al., 2003) becomes both inefficient and time consuming due to the presence of numerical integration in the cost function.

This paper uses the approach presented in (Farahani et al., 2010), to approximate the calculation of the integrals using raw moments of random variables. This method simplifies the computations considerably and by choosing the appropriate order p of the raw moments, the approximation error can be made sufficiently small. Since we can compute these moments analytically, this approach results in a much faster and more efficient way to solve the identification problem for stochastic MPL systems without increasing the computational complexity.

The paper is organized as follows. Section 2 presents a brief description of the max-plus algebra and stochastic MPL systems. In Section 3 the identification of stochastic MPL systems is considered, as presented in (van den Boom et al., 2003). Section 4 introduces the proposed approximation method, which is based on the higher-order raw moments. In this section it is also shown how this approach reduces the complexity of the identification problem. Section 5 presents a worked example in which the new method is applied and the computation time of

the approximation method is compared with the one in (van den Boom et al., 2003).

2. MAX-PLUS ALGEBRA AND STOCHASTIC MAX-PLUS LINEAR SYSTEMS

In this section we concisely present the basics of the max-plus algebra and the class of stochastic max-plus linear systems. More detailed information can be found in (Baccelli et al., 1992; Cuninghame-Green, 1979; Heidergott et al., 2006).

2.1 Max-Plus Algebra

In the max-plus algebra, the main operations are maximization and addition, defined as follows:
$$x \oplus y = \max(x, y)$$
$$x \otimes y = x + y$$
for $x, y \in \mathbb{R}_\varepsilon$ where $\mathbb{R}_\varepsilon = \mathbb{R} \cup \{\varepsilon\}$ and $\varepsilon = -\infty$. Note that the zero element of the max-plus addition is ε, i.e., $x \oplus \varepsilon = x$, and the identity element of the max-plus multiplication is $e = 0$, i.e., $x \otimes e = x$. The corresponding max-plus matrix operations are defined (Baccelli et al., 1992) as
$$(A \oplus B)_{ij} = a_{ij} \oplus b_{ij} = \max(a_{ij}, b_{ij})$$
$$(A \otimes C)_{ij} = \bigoplus_{k=1}^{n} a_{ik} \otimes c_{kj} = \max_{k=1,\dots,n} (a_{ik} + c_{kj})$$
for $A, B \in \mathbb{R}_\varepsilon^{m \times n}$ and $C \in \mathbb{R}_\varepsilon^{n \times p}$. In the sequel of this paper, we denote the i-th row of matrix A by $A_{i,\cdot}$ and the j-th column by $A_{\cdot,j}$. To avoid ambiguity, wherever we have a multiplication in conventional algebra, we drop the multiplication sign, while in the max-plus expressions we always keep the sign $\otimes$ present.

2.2 Max-Plus-Nonnegative-Scaling Functions

Denote the set of max-plus-nonnegative-scaling functions by $\mathscr{S}_{\mathrm{mpns}}$. It consists of functions f of the form
$$f(z) = \max_{i=1,\dots,m} (\tau_{i,1} z_1 + \cdots + \tau_{i,n} z_n + \xi_i)$$
with variable $z \in \mathbb{R}_\varepsilon^n$ and constant coefficients $\tau_{i,j} \in \mathbb{R}^+$ and $\xi_i \in \mathbb{R}$, where $\mathbb{R}^+$ is the set of the nonnegative real numbers. In the sequel, we stress that f is a function of z by writing $f \in \mathscr{S}_{\mathrm{mpns}}(z)$. As shown by van den Boom and De Schutter (2004), the set $\mathscr{S}_{\mathrm{mpns}}$ is closed under the operations $\oplus, \otimes$, and the scalar multiplication by a nonnegative scalar.

2.3 Stochastic MPL Systems

Discrete event systems with synchronization but no choice can be modeled as follows (Baccelli et al., 1992; Cuninghame-Green, 1979):
$$x(k+1) = A(k) \otimes x(k) \oplus B(k) \otimes u(k) \tag{1}$$
$$= [A(k)\ B(k)] \otimes \begin{bmatrix} x(k) \\ u(k) \end{bmatrix} \tag{2}$$
$$= Q(k) \otimes \phi(k) \tag{3}$$
where $x(k)$ is the state of the system at event step k, and $u(k)$ is the input of the system or in fact the time instant at which the internal input event occurs for the k-th time. In a stochastic system noise and modeling errors lead to perturbation of the system matrices, $A(k)$ and $B(k)$. Following van den Boom et al. (2003), we present these uncertainties in a single framework,

using one stochastic vector $e(k)$ with a certain probability distribution. Note that the entries of the system matrices belong to $\mathscr{S}_{\mathrm{mpns}}(e(k))$ (van den Boom and De Schutter, 2004), i.e., $A(k) \in \mathscr{S}_{\mathrm{mpns}}^{n \times n}(e(k))$, $B(k) \in \mathscr{S}_{\mathrm{mpns}}^{n \times n_u}(e(k))$. In the sequel, we denote the uncertain system matrices with matrix $Q(k)$ and the state and input vector with $\phi(k)$ (cf. (3)).

In order to identify the unknown system parameters, we need to distinguish between the a priori known parameters, i.e., the parameters that are either constant or determined in advance such as the transportation time in a production system, and ones that have to be identified. Therefore, the i-th row of the matrix $Q(k)$ can be written as:
$$Q_{i,\cdot}(k) = \Xi_{i,\cdot} + \theta^T \Delta^{(i)} + e^T(k) \Lambda S^{(i)} \tag{4}$$
where Ξ represents the parameters that are known, θ is a vector of unknown parameters, $e(k) = [e_1(k), \dots, e_{n_e}(k)]^T$ is the stochastic vector, the diagonal matrix $\Lambda = \mathrm{diag}(\lambda_1, \dots, \lambda_{n_e})$ contains the amplitude of the noise, and $\Delta^{(i)}$ and $S^{(i)}$ are selection matrices for the i-th row with zeros and ones as entries. In fact the role of the selection matrices is to determine which elements of the vectors θ and $e(k)$ have to appear in the i-th row. We assume that the probability density function of $e(k)$ (denoted by $f(e)$) and other matrices are known a priori and the only parameters that have to be identified are the components of θ and the diagonal elements of Λ, denoted by $\lambda = [\lambda_1, \dots, \lambda_{n_e}]^T$. Therefore, in the identification procedure we will derive estimates $\hat{\theta}$ and $\hat{\lambda}$ of θ and λ, respectively.

3. IDENTIFICATION OF STOCHASTIC MAX-PLUS LINEAR SYSTEMS

In this paper we perform the identification based on input-state data. Note that in MPL systems the state contains the time instants at which the state events occur. Since the state is observable by assumption, these instants can be measured easily and so we usually have full-state information. Consider the input-state sequence $\{(u(k), x(k)\}_{k=1}^{N}$ of a system of the form (3) and assume that the system parameters $\hat{\theta}$ and $\hat{\lambda}$ has to be identified using this sequence. Following van den Boom et al. (2003) we assume that the input-state sequence is sufficiently rich [1] to capture all the relevant information about the system (see also Schullerus and Krebs (2001b); Schullerus et al. (2003)). We consider the following identification problem (van den Boom et al., 2003)
$$(\theta^*, \lambda^*) = \arg \min_{(\hat{\theta}, \hat{\lambda})} J(\hat{\theta}, \hat{\lambda}) \tag{5}$$
$$\text{subject to} \qquad \hat{\lambda} > 0$$
while
$$J(\hat{\theta}, \hat{\lambda}) = \sum_{k=1}^{N-1} \sum_{i=1}^{n} (\mathbb{E}[x_i(k+1|k)] - x_i(k+1))^2 \tag{6}$$
where $\mathbb{E}[\cdot]$ denotes the expected value operator and $\mathbb{E}[x_i(k+1|k)]$ is the one-step-ahead prediction of x_i for event step $k+1$,

---

[1] Intuitively, this can be characterized as follows. Note that (3) and (4) imply that each component of $x(k+1)$ can be written as a max expression of terms in which the unknown parameters θ and λ appear. An input signal is then said to be sufficiently rich if it is such that each of these terms is the maximum one sufficiently often (this is also related to the idea of persistent excitation in conventional system identification (Ljung, 1999)).

using the knowledge from event step k. Considering (3) and (4), we can rewrite the one-step-ahead prediction as

$$\mathbb{E}[x_i(k+1|k)] = \mathbb{E}\big[\big(\Xi_{i,\cdot} + \theta^T\Delta^{(i)} + e^T(k)\Lambda S^{(i)}\big) \otimes \phi(k)\big]$$

and hence, the one-step-ahead prediction error is

$$\hat{\eta}_i(k+1,\hat{\theta},\hat{\lambda}) = \mathbb{E}[x_i(k+1|k)] - x_i(k+1)$$
$$= \mathbb{E}\big[\max_{j=1,\dots,m}\big(\xi_{i,j} + \hat{\theta}^T\Delta^{(i)}_{\cdot,j} + e^T(k)\hat{\Lambda}S^{(i)}_{\cdot,j}$$
$$+ \phi_j(k) - x_i(k+1)\big)\big] \quad (7)$$

Now for a specific realization of the noise vector $e(k)$, let:

$$\eta_i(k+1,\hat{\theta},\hat{\lambda},e(k)) = \max_{j=1,\dots,m}\big(\xi_{i,j} + \hat{\theta}^T\Delta^{(i)}_{\cdot,j} + e^T(k)\hat{\Lambda}S^{(i)}_{\cdot,j}$$
$$+ \phi_j(k) - x_i(k+1)\big)$$

Hence,

$$\hat{\eta}_i(k+1,\hat{\theta},\hat{\lambda}) = \mathbb{E}[\eta_i(k+1,\hat{\theta},\hat{\lambda},e(k))]$$

To have a more compact notation, let $\alpha_{ij}(k) = \xi_{i,j} + \phi_j(k) - x_i(k+1)$, $\Pi_{ij} = \Delta^{(i)}_{\cdot,j}$, and $\Gamma_{ij} = \text{diag}(S^{(i)}_{1,j},\dots,S^{(i)}_{n_e,j})$. Since $e^T(k)\hat{\Lambda}S^{(i)}_{\cdot,j}$ is a scalar and $\hat{\Lambda}$ is a diagonal matrix, we have:

$$e^T(k)\hat{\Lambda}S^{(i)}_{\cdot,j} = (S^{(i)}_{\cdot,j})^T\hat{\Lambda}e(k) = \hat{\lambda}^T\Gamma_{ij}e(k)$$

Therefore, we can rewrite $\eta_i(k+1,\hat{\theta},\hat{\lambda},e(k))$ as

$$\eta_i(k+1,\hat{\theta},\hat{\lambda},e(k)) = \max_{j=1,\dots,m}\big(\alpha_{ij} + \Pi^T_{ij}\hat{\theta} + \hat{\lambda}^T\Gamma_{ij}e(k)\big) \quad (8)$$

Therefore, the expected value of $\eta_i(k+1,\hat{\theta},\hat{\lambda},e(k))$, i.e., $\hat{\eta}_i(k+1,\hat{\theta},\hat{\lambda})$ can be computed analytically as follows:

$$\hat{\eta}_i(k+1,\hat{\theta},\hat{\lambda}) = \int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty}\eta_i(k+1,\hat{\theta},\hat{\lambda},e(k))f(e)\mathrm{d}e$$
$$= \sum_{j=1}^{n_v}\int\cdots\int_{e\in\Omega_{ij}(\hat{\theta},\hat{\lambda},k)}\big(\alpha_{ij} + \Pi^T_{ij}\hat{\theta} + \hat{\lambda}^T\Gamma_{ij}e(k)\big)f(e)\mathrm{d}e$$
$$(9)$$

where $\mathrm{d}e = \mathrm{d}e_1,\dots,\mathrm{d}e_{n_e}$ and the sets $\Omega_{ij}(\hat{\theta},\hat{\lambda},k), i = 1,\dots,n$, $j = 1,\dots,m$ are defined such that they have non-overlapping interiors, and that for all $e \in \Omega_{ij}(\hat{\theta},\hat{\lambda},k)$,

$$\eta_i(k+1,\hat{\theta},\hat{\lambda},e(k)) = \alpha_{ij} + \Pi^T_{ij}\hat{\theta} + \hat{\lambda}^T\Gamma_{ij}e(k)$$

and for any i it holds that $\bigcup_{j=1}^m\Omega_{ij}(\hat{\theta},\hat{\lambda},k) = \mathbb{R}^{n_e}$, i.e., for all realizations of e, the j-th term in (8) gives the maximum, and the sets $\Omega_{ij}(\hat{\theta},\hat{\lambda},k)$ cover the whole space of $\mathbb{R}^{n_e}$ and only overlap at the boundary of the regions.

4. APPROXIMATION METHOD

Due to the numerical integration presented in (9), the computational complexity of the identification problem grows fast when the number of stochastic variables increases. Also, in the extension to multi-step ahead prediction (van den Boom et al., 2003), the stochastic complexity grows drastically. van den Boom et al. (2003) suggested to use the piecewise polynomial probability density functions as an approximation of the real probability density function. This method leads to a two step procedure: In the first step all the regions on which each term in (9) is maximum, has to be specified. In the second step the integral over these regions has to be calculated. Therefore, the whole procedure is quite complex and time consuming. Accordingly, we propose an approximation method, based on the p-th raw moments of a stochastic variable, as explained in (Farahani et al., 2010). We consider the p-norm and the ∞-norm

of a random vector $y = [y_1,\dots,y_n]^T$, i.e., $\|y\|_p = \big(|y_1|^p + \cdots + |y_n|^p\big)^{1/p}$ and $\|y\|_\infty = \max(|y_1|,\dots,|y_n|)$ respectively, and the relation between these two norms (Golub and Van Loan, 1990), i.e.,

$$\|y\|_\infty \leq \|y\|_p \leq n^{1/p}\|y\|_\infty \quad (10)$$

However, in contrast to the definition of p-norm and ∞-norm, the definition of (8) has no absolute value. Clearly, $y \leq |y|$ and y has an unbounded distribution in general. We introduce a new variable L as an approximate lower bound, i.e., $L \leq y_j$ for $j = 1,\dots,n$. Hence,

$$\max(y_1,\dots,y_n) = \max(y_1 - L,\dots,y_n - L) + L$$
$$\leq \max(|y_1 - L|,\dots,|y_n - L|) + L. \quad (11)$$

Hence, the role of L is to decrease the error of approximating $y - L$ by $|y - L|$.

Now assume that the elements of vector y are independent and identically normally distributed, i.e., $y_j \sim \mathcal{N}(\mu_j,\sigma_j^2)$. According to the "$3\sigma$-rule", 99.7% of the observations of a normally distributed random variable fall within the mean minus 3 times the standard deviation and the mean plus 3 times the standard deviation. Therefore, we choose $L = \min_{j=1,\dots,n}(\mu_j - 3\sigma_j)$ as an approximate lower bound. There is an important theorem that must be mentioned before we proceed further.

Theorem 1. (Jensen's Inequality Boyd and Vandenberghe (2004)). Let x be an integrable real-valued random variable and φ a concave function such that $\varphi(x)$ is integrable. Then: $\varphi(\mathbb{E}[x]) \geq \mathbb{E}[\varphi(x)]$.

Now let $x_j = y_j - L$ and by considering (10), (11), the Jensen's inequality for concave functions, and noting the fact that expected value is a monotonic operator, we obtain:

$$\mathbb{E}[\max(x_1,\dots,x_n)] + L \leq \left(\sum_{j=1}^n \mathbb{E}[|x_j|^p]\right)^{1/p} + L \quad (12)$$

Hence, we can use the left-hand side of (12) as an approximation of the right-hand side. Consequently, we can approximate the function $\hat{\eta}_i(k+1,\hat{\theta},\hat{\lambda})$ by $\hat{\eta}_{\text{app},i}(\hat{\theta},\hat{\lambda})$ for an appropriate choice of p where

$$\hat{\eta}_{\text{app},i}(\hat{\theta},\hat{\lambda}) = \left(\sum_{j=1}^m \mathbb{E}\big[(\alpha_{ij} + \Pi^T_{ij}\hat{\theta} + \hat{\lambda}^T\Gamma_{ij}e(k) - L)^p\big]\right)^{1/p}$$
$$+ L \quad (13)$$

For an even positive integer $p = 2q$, $q \in \mathbb{Z}^+$, $\mathbb{E}[x^p] = \mathbb{E}[|x|^p]$. Therefore in the rest of this paper, we use $\mathbb{E}[x^p]$ where p is an even integer larger than or equal to 2.

Following (Farahani et al., 2010), we also assume that the elements of the stochastic vector e are independent and identically normally distributed, i.e., $e_\ell \sim \mathcal{N}(\mu_\ell,\sigma_\ell^2)$ for $\ell = 1,\dots,n_e$. By using the property of the normal distribution that sum of the independent normally distributed random variables has also a normal distribution (Dekking et al., 2005), we conclude that the random variable $\omega_j = \alpha_{ij} + \Pi^T_{ij}\hat{\theta} + \hat{\lambda}^T\Gamma_{ij}e(k) - L$ in (13) is also normally distributed with mean $\mu_j = \alpha_{ij} + \Pi^T_{ij}\hat{\theta} + \hat{\lambda}^T\Gamma_{ij}\mu_e - L$ and variance $\sigma_j^2 = \sum_{\ell=1}^{n_e}(\hat{\lambda}^T\Gamma_{ij})_\ell^2\sigma_\ell^2$. Note that $L =$

$\min\limits_{j=1,\dots,m} (\alpha_{ij}+\Pi_{ij}^T\hat\theta+\hat\lambda^T\Gamma_{ij}\mu_e-3\sigma_j)$ and $\mu_e=[\mu_1,\dots,\mu_{n_e}]^T$ is the mean of the error vector e.

Since we know that the random variable $\omega_j = \alpha_{ij}+\Pi_{ij}^T\hat\theta+\hat\lambda^T\Gamma_{ij}e(k)-L$ is normally distributed, we can define its p-th raw moments. By definition, the p-th raw moment of a normally distributed random variable x with mean μ and standard deviation σ can be computed as follows:

$$\mathbb{E}[x^p] = \int_{-\infty}^{\infty} x^p \frac{1}{\sqrt{2\pi}\sigma} e^{-(x-\mu)^2/(2\sigma^2)}\mathrm{d}x$$

which is finite for all integers $p \geq 2$. According to (Willink, 2005), this moment has a closed form which can be expressed as follows:

$$\mathbb{E}\left[x^p\right] = \sigma^p i^{-p} H_p(i\mu/\sigma) \qquad (14)$$

where

$$H_p(x) \equiv (-1)^p \exp(x^2/2)\frac{d^p}{dx^p}\exp(-x^2/2)$$
$$= p!\sum_{k=0}^{p/2}\frac{(-1)^k x^{p-2k}}{2^k k!(p-2k)!}$$

is the p-th order Hermite polynomial, and where the second equality can be obtained by considering equations (26.2.51) and (22.3.11) in (Abramowitz and Stegun, 1964). Note that the right-hand side of (14) is real because $H_p(x)$ contains only even powers of x if p is even. Therefore, by using (14), we can rewrite the approximation function (13) as follows:

$$\hat\eta_{\mathrm{app},i}(\hat\theta,\hat\lambda) = \left(\sum_{j=1}^m \sigma_j^p i^{-p} H_p(i\mu_j/\sigma_j)\right)^{1/p} + L.$$

Furthermore, we can obtain its gradients with respect to $\hat\theta$ and $\hat\lambda$ by considering the recurrence relations for Hermite polynomials (Abramowitz and Stegun, 1964), i.e.,

$$H_{p+1}(x) = xH_p(x) - pH_{p-1}(x)$$

and

$$H'_p(x) = pH_{p-1}(x),$$

Therefore:

$$\nabla_{\hat\theta}\hat\eta_{\mathrm{app},i}(\hat\theta,\hat\lambda) = \left(\sum_{j=1}^m \sigma_j^p i^{-p} H_p(i\mu_j/\sigma_j)\right)^{1/p-1}$$
$$\cdot\left(\sum_{j=1}^m \beta_{ij}\sigma_j^{p-1} i^{-p+1} H_{p-1}(i\mu_j/\sigma_j)\right)$$

and

$$\nabla_{\hat\lambda}\hat\eta_{\mathrm{app},i}(\hat\theta,\hat\lambda) = \left(\sum_{j=1}^m \sigma_j^p i^{-p} H_p(i\mu_j/\sigma_j)\right)^{1/p-1}$$
$$\cdot\Bigg(\sum_{j=1}^m \sigma_j^{p-2} i^{-p}\big[\gamma_{ij}\sigma_e\sigma_j H_p(i\mu_j/\sigma_j)$$
$$+ (\gamma_{ij}\mu_e\sigma_j - \gamma_{ij}\sigma_e\mu_j)iH_{p-1}(i\mu_j/\sigma_j)\big]\Bigg)$$

where μ_j and σ_j are the same as defined before.

As a result, we can approximate $J(\hat\theta,\hat\lambda)$ in (6) by replacing $\hat\eta_i(k+1,\hat\theta,\hat\lambda)$ with $\hat\eta_{\mathrm{app},i}(\hat\theta,\hat\lambda)$ as follows:

$$J_{\mathrm{app}}(\hat\theta,\hat\lambda) = \sum_{k=1}^N \sum_{i=1}^n \left(\hat\eta_{\mathrm{app},i}(\hat\theta,\hat\lambda)\right)^2 \qquad (15)$$

with the subgradients

$$\nabla_{\hat\theta}J_{\mathrm{app}}(\hat\theta,\hat\lambda)=\sum_{k=1}^N \sum_{i=1}^n 2\hat\eta_{\mathrm{app},i}(\hat\theta,\hat\lambda)\nabla_{\hat\theta}\hat\eta_{\mathrm{app},i}(\hat\theta,\hat\lambda)$$

$$\nabla_{\hat\lambda}J_{\mathrm{app}}(\hat\theta,\hat\lambda)=\sum_{k=1}^N \sum_{i=1}^n 2\hat\eta_{\mathrm{app},i}(\hat\theta,\hat\lambda)\nabla_{\hat\lambda}\hat\eta_{\mathrm{app},i}(\hat\theta,\hat\lambda)$$

and solve the optimization problem by means of a gradient search method.

5. EXAMPLE

In this example, we apply the proposed approximation method to the identification problem of a stochastic max-plus linear system to check whether it works efficiently and gives reliable results. Moreover, we can compare the computation time of this method with the one from the analytic solution in (van den Boom et al., 2003).

Consider the max-plus-linear system (1) with the following system matrices

$$A(k) = \begin{bmatrix} \theta_1(k) & \theta_2(k) \\ \varepsilon & \varepsilon \end{bmatrix}$$
$$B(k) = \begin{bmatrix} \theta_3(k) \\ 0 \end{bmatrix}$$

where the true parameter vector θ is given by

$$\theta = [\,\theta_1\ \theta_2\ \theta_3\,]^T = [\,0.05\ 1\ 0.5\,]^T$$

These parameters are perturbed by noise components $e_\ell(k)$ that are independent and have a standard normal distribution, i.e., $e_\ell(k) \sim \mathcal{N}(0,1)$, for $\ell = 1,\dots,3$, and with amplitudes

$$\lambda = [\,\lambda_1\ \lambda_2\ \lambda_3\,]^T = [\,0.3\ 0.3\ 0.3\,]^T.$$

In this simulation study we simulate the system for 400 event steps, i.e., for k=1,...,400. A parameter estimation is done with the input-state data where the input signal is a staircase signal with an average slope of 1.5, given by

$$u(k) = 4.5\cdot\left(1+\lfloor k/3\rfloor\right)$$

where $\lfloor x\rfloor$ denotes the largest integer less than or equal to x. The input signal $u(k)$ in shown in Figure 1 for $k = 1,\dots,40$.

We minimize the approximate cost function (15) based on the one-step ahead prediction, i.e., we predict the behavior of the system at the event step $k+1$ based on the information that we have at the event step k. Also we use a gradient-based, multi-start, optimization method, i.e., sequential quadratic programming (SQP) method, considering 15 different initial values to start the optimization with, reporting the minimum of the optimal values as the estimated parameter. By means of experiments, we have found that $p = 30$ gives a good approximation in this specific example. As a result of the multi-start optimization, we obtained the estimated parameter vector

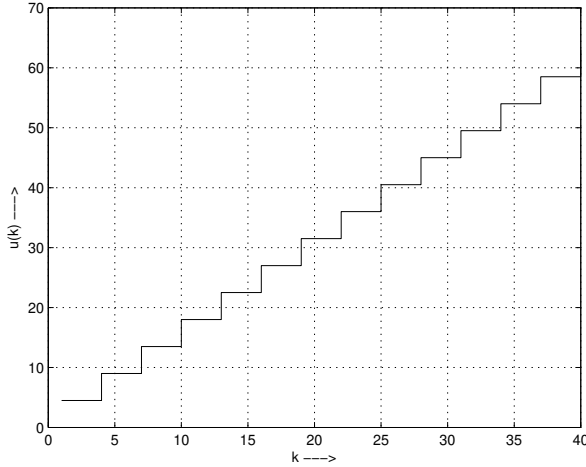$$\hat\theta = [0.0400\ \ 0.9969\ \ 0.4946]^T$$

Fig. 1. The input signal $u(k)$

and the estimated noise amplitude vector

$$\hat{\lambda} = [0.0970 \quad 0.0480 \quad 0.0502]^T.$$

The quality of these results is comparable to the ones of (van den Boom et al., 2003). Note that in this example, due to presence of the standard normal noise, the influence of the noise on the system is larger than in the case of (van den Boom et al., 2003), where the noise has a uniform distribution. This phenomenon and the error caused by Jensen's inequality in the approximation method both lead to a less satisfactory estimation for the noise amplitude vector, $\hat{\lambda}$.

Another issue that can be discussed here is the computation time. One of the goals of using the proposed approximation method (13) was to decrease the computation time. For 400 event steps, the computation time [2] of this example (with the presence of a normally distributed noise sequence) is 569 s, which is much less than the time obtained using the method of (van den Boom et al., 2003), which is 4277 s, for the same number of event steps. It is important to note that the computation time that we obtained is for the case that the error vector is normally distributed while in (van den Boom et al., 2003) the error vector is uniformly distributed, which is the simplest one among the distributions due to the constant probability density function which results in easy integrals, and yet our computation time is much less than the one in (van den Boom et al., 2003). If the method of (van den Boom et al., 2003) is applied to the case with the normally distributed error vector (which involves multi-dimensional numerical integration over exponential instead of constant functions), the computation time would be even much higher than 4277 s. Therefore, the new method increases the time efficiency while still guaranteeing a similar performance to .

6. CONCLUSIONS

We have proposed the application of an approximation method based on the p-th raw moment of a random variable to the identification problem for stochastic max-plus linear systems. This approximation method decreases the computational complexity considerably (especially for large systems) since the numerical integration involved in the solution of the identification

problem is approximated by an analytic solution. Although we assumed that the elements of the noise vector in the system are normally distributed, this approximation method is applicable to any distribution that is preserved under the summation and a closed form of its p-th raw moment exists, such as Poisson and Gamma distributions (Papoulis, 1991). Moreover, since an explicit expression for the gradient of the criterion function can be calculated, the parameter estimation can be done using a gradient-based optimization method.

One topic for future research is the development of algorithms for stochastic max-plus linear systems based on input-output data (instead of input-state data) or with only partial state information, based on the proposed approximation method. Another interesting topic is to study the exact effect of Jensen's inequality on the estimation of the noise amplitude, and to find a method to specify the most appropriate moment order p, in order to obtain a better estimation. Yet another topic would be to check under which condition the identification method converges to the true solution, specifically when the proposed approximation method is applied to it. It is also interesting to extend the proposed approach to other classes of discrete-event systems such as max-min-plus systems.

REFERENCES

Abramowitz, M.A. and Stegun, I. (1964). *Handbook of Mathematical Functions*. National Bureau of Standards, US Government Printing Office, Washington DC.

Baccelli, F., Cohen, G., Olsder, G., and Quadrat, J. (1992). *Synchronization and Linearity*. John Wiley & Sons, New York.

Boyd, S. and Vandenberghe, L. (2004). *Convex Optimization*. Cambridge University Press, London, UK.

Cassandras, C. and Lafortune, S. (1999). *Introduction to Discrete Event Systems*. Kluwer Academic Publishers, Boston, Massachusetts.

Cuninghame-Green, R.A. (1979). *Minimax Algebra*, volume 166 of *Lecture Notes in Economics and Mathematical Systems*. Springer-Verlag, Berlin, Germany.

De Schutter, B., van den Boom, T.J.J., and Verdult, V. (2002). State space identification of max-plus-linear discrete event systems from input-output data. In *Proceedings of the 41st IEEE Conference on Decision and Control*, 4024–4029. Las Vegas, Nevada.

Dekking, F., Kraaikamp, C., Lopuhaä, H.P., and Meester, L.E. (2005). *A Modern Introduction to Probability and Statistics*. Springer, London.

Farahani, S.S., van den Boom, T., van der Weide, H., and De Schutter, B. (2010). An approximation approach for model predictive control of stochastic max-plus linear systems. In *Proceedings of the workshop on Discrete Event Systems (WODES)*, 386–391. Berlin, Germany.

Golub, G.H. and Van Loan, C.F. (1990). *Matrix Computations*. The John Hopkins University Press, Baltimore, Maryland.

---

[2] These times were obtained running Matlab 7.5.0 (R2007b) on a 2.33 GHz Intel Core Duo E655 processor.

Heidergott, B., Olsder, G.J., and van der Woude, J. (2006). *Max Plus at Work*. Princeton University Press, Princeton, New Jersey.

Ho, Y. (ed.) (1992). *Discrete Event Dynamic Systems: Analyzing Complexity and Performance in the Modern World*. IEEE Press, Piscataway, New Jersey.

Ljung, L. (1999). *System Identification: Theory for the User*. Prentice-Hall, Upper Saddle River, New Jersey, 2nd edition.

Mairesse, J. (1994). Stochastic linear systems in the (max,+) algebra. *Springer*, 199, 257–265.

Olsder, G.J., Resing, J.A.C., de Vries, R.E., Keane, M., and Hooghiemstra, G. (1990). Discrete event systems with stochastic processing times. *IEEE Transactions on Automatic Control*, 35(3), 299–302.

Papoulis, A. (1991). *Probability, Random Variables, and Stochastic Processes*. McGraw-Hill, Singapore.

Peterson, J. (1981). *Petri Net Theory and the Modeling of Systems*. Prentice-Hall, Englewood Cliffs, New Jersey.

Resing, J.A.C., de Vries, R.E., Hooghiemstra, G., Keane, M.S., and Olsder, G.J. (1990). Asymptotic behavior of random discrete event systems. *Stochastic Processes and Their Applications*, 36(2), 195–216.

Schullerus, G. and Krebs, V. (2001a). Diagnosis of batch processes based on parameter estimation of discrete event models. In *Proceedings of the European Control Conference 2001 (ECC'01)*, 1612–1617. Porto, Portugal.

Schullerus, G. and Krebs, V. (2001b). Input signal design for discrete event model based batch process diagnosis. In *Proceedings of the 4th IFAC Workshop on On-Line Fault Detection and Supervision in the Chemical Process Industries*. Chejudo Island.

Schullerus, G., Krebs, V., De Schutter, B., and van den Boom, T. (2003). Optimal input signal design for identification of max-plus-linear systems. In *Proceedings of the 2003 European Control Conference (ECC'03)*. Cambridge, UK. Paper 026.

van den Boom, T.J.J. and De Schutter, B. (2004). Model predictive control for perturbed max-plus-linear systems: A stochastic approach. *International Journal of Control*, 77(3), 302–309.

van den Boom, T., De Schutter, B., and Verdult, V. (2003). Identification of stochastic max-plus-linear systems. In *Proceedings of the 2003 European Control Conference (ECC'03)*. Cambridge, UK. Paper 104.

Willink, R. (2005). Normal moments and Hermite polynomials. *Statistics & Probability Letters*, 73(3), 271–275.