# Identification of Stochastic Max-Plus-Linear Systems

T.J.J. van den Boom, B. De Schutter and V. Verdult

Delft Center for Systems and Control
Delft University of Technology, Mekelweg 2, 2628 CD Delft, The Netherlands
Phone: +31-15-2784052/5113, Fax: +31-15-2786679
email: {t.j.j.vandenboom,b.deschutter,v.verdult}@dcsc.tudelft.nl

**Abstract**

We present a method to identify the parameters of a state space model for a max-plus-linear discrete event system from measured data. Previous papers report on results with noise-free measured data. In this paper we extend this to identification for perturbed max-plus-linear systems in a stochastic setting. The approach is based on recasting the identification problem as an optimization problem. We show that under quite general conditions the resulting optimization problems can be solved very efficiently using gradient search techniques.

**Keywords**: discrete event systems, system identification, max-plus-linear systems, noise, stochastic setting.

## 1 Introduction

Discrete event systems (DES) are dynamic, asynchronous systems the state of which changes due to the occurrence of events; this in contrast to continuous variable systems, whose behavior is governed by the progression of time or the ticks of a clock and which can be modeled by a system of differential or difference equations. Typical examples of DES are flexible manufacturing systems, telecommunication networks, parallel processing systems, traffic control systems, and logistic systems. An event corresponds to the start or the end of an activity. In the case of a production system possible events are: the completion of a part on a machine, a machine breakdown, or a buffer becoming empty. There exist many modeling frameworks for DES such as queueing theory, (extended) state machines, formal languages, automata, temporal logic models, generalized semi-Markov processes, Petri nets, and computer simulation models (see [4, 10, 14] and the references therein). In general, models that describe the behavior of a DES are nonlinear in conventional algebra. However, there is a class of DES that can be described by a model that is "linear" in the max-plus algebra [1, 5]. Such DES are called max-plus-linear (MPL) DES. Essentially, they can be characterized as DES in which only synchronization and no concurrency or choice occurs. So typical examples are serial production lines, production systems with a fixed routing schedule, and railway networks.

If we want to use a model for control and other purposes, we have to be able to determine the parameters of the model. Most identification methods for MPL DES use a transfer function approach [2, 8]. One could argue that an identified transfer function description can be transformed into a state space model, but then the connection with the physical structure is usually lost[1]. Furthermore, 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. In addition, the Model Predictive Control framework of [6] requires a state space model. Therefore, we focus on state space identification for MPL DES.

In [7, 16, 17, 18] state space identification methods have been derived in which the internal structure of the system is assumed to be completely known and the state is assumed to be measurable. In [7] we assumed that only input-output sequences were available. The method can also be used for fully parameterized state space identification and also yields an estimate of the state sequence. All these methods use noise free data.

In this paper we look at the noisy case. In contrast to conventional linear systems, where noise and disturbances are usually modeled by including an extra term in the system equations (i.e., the noise is considered to be additive), the influence of noise and disturbances in MPL DES is not max-plus-additive, but max-plus-multiplicative (see [1]). This means that the system parameters will be perturbed and as a consequence the system identification has to take the stochastic properties into account. We will show that under quite general conditions the resulting identification problem can be solved very efficiently.

This paper is organized as follows. In Section 2 we first give a concise introduction to MPL DES in a stochastic framework. In Section 3 we derive an expression for the prediction error in the stochastic setting, and we suggest an iden-

[1] For MPL DES the physical layout of the system is often clearly recognizable in the structure of the state space matrices (see also Section 5).

tification algorithm. Section 4 discusses the computational aspects of the algorithm and in Section 5 we give a worked example.

## 2 Stochastic max-plus-linear DES

In this section we define the class of stochastic max-plus-linear DES. For this purpose we will first give the basic definition of the max-plus algebra and min-plus algebra.

**Max-plus algebra**
Define $\varepsilon = -\infty$ and $\mathbb{R}_\varepsilon = \mathbb{R} \cup \{\varepsilon\}$. The max-plus-algebraic addition ( $\oplus$ ) and multiplication ($\otimes$) are defined as follows [1, 5]:

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

for numbers $x, y \in \mathbb{R}_\varepsilon$, and

$$[P \oplus Q]_{i,j} = P_{ij} \oplus Q_{i,j} = \max(P_{i,j}, Q_{i,j})$$
$$[P \otimes R]_{i,j} = \bigoplus_{k=1}^{n} P_{i,k} \otimes R_{k,j} = \max_{k=1,\ldots,n} (P_{i,k} + R_{k,j})$$

for matrices $P, Q \in \mathbb{R}_\varepsilon^{m \times n}$ and $R \in \mathbb{R}_\varepsilon^{n \times p}$ and where $P_{i,j}$ is the $ij$th element of matrix $P$. In the sequel of the paper we will use the notation $P_{i,\cdot}$ for the $i$th row of matrix $P$ and $P_{\cdot,j}$ for the $j$th column of matrix $P$.

Note that in this paper we use both max-plus and conventional algebra. Therefore we will always write $\oplus$ and $\otimes$ explicitly in all equations. The operations '$+$' and '$\cdot$' denote the conventional summation and multiplication operators (Only the conventional multiplication operator is sometimes omitted).

**Max-plus-linear systems**
In [1] a state-space setting is used to DES in which there is synchronization but no concurrency, described by the state equation

$$\begin{aligned} x(k+1) &= A \otimes x(k) \oplus B \otimes u(k) && (1)\\ &= \begin{bmatrix} A & B \end{bmatrix} \otimes \begin{bmatrix} x(k) \\ u(k) \end{bmatrix} && (2)\\ &= \Theta \otimes p(k) && (3) \end{aligned}$$

where

$$\Theta = \begin{bmatrix} A & B \end{bmatrix} \in \mathbb{R}_\varepsilon^{n \times m}, \qquad p(k) = \begin{bmatrix} x(k) \\ u(k) \end{bmatrix} \in \mathbb{R}_\varepsilon^{m}.$$

DES that can be described by this model will be called max-plus-linear (MPL) systems. The index $k$ is called the event counter. The input $u(k)$ contains the time instants at which the input events occur for the $k$th time, and the state $x(k)$ contains the time instants at which the state events occur for the $k$th time[2].

[2] More specifically, for a manufacturing system, $x(k)$ contains the time instants at which the processing units start working for the $k$th time, and $u(k)$ contains the time instants at which the $k$th batch of raw material is fed to the system.

**Stochastic max-plus-linear systems**
In this paper we consider systems of the form (1), but in a stochastic setting, as described in [20]. In contrast to conventional linear systems, where noise and disturbances are usually modeled by including an extra term in the system equations (i.e., the noise is considered to be additive), the influence of noise and disturbances in MPL systems is not max-plus-additive, but max-plus-multiplicative. This means that the system parameters will be perturbed and as a consequence the system properties will change.

Consider the following MPL system

$$\begin{aligned} x(k+1) &= A(k) \otimes x(k) \oplus B(k) \otimes u(k) && (4)\\ &= \begin{bmatrix} A(k) & B(k) \end{bmatrix} \otimes \begin{bmatrix} x(k) \\ u(k) \end{bmatrix} && (5)\\ &= \Theta(k) \otimes p(k) && (6) \end{aligned}$$

where $\Theta(k)$ represents the uncertain system matrix due to disturbances and noise. The uncertainty caused by disturbances, is gathered in the uncertainty vector $e(k) \in \mathbb{R}^{n_e}$. In this paper we assume that the uncertainty has stochastic properties. Hence, $e(k)$ is a stochastic variable. To make the identification of the unknown system parameters possible, we have to distinguish the parameters of $\Theta$ that are known a priori (because of the known structure), the parameters of $\Theta$ that we like to identify, and an additional term to account for the noise and disturbances. We therefore assume that the $i$th row of system matrix $\Theta(k)$ can be written as:

$$\Theta_{i,\cdot}(k) = R_{i,\cdot} + \theta^T T^{(i)} + e(k)^T \Lambda S^{(i)} \qquad (7)$$

where $R$ represent the parameters that are known, $\theta$ is a vector with all unknown parameters of $\Theta$, the diagonal matrix $\Lambda$ contains the amplitude of the noise, and $T^{(i)}$ and $S^{(i)}$ are selection matrices for the $i$th row with only ones and zeros. We assume that the selection matrices $T^{(i)}$ and $S^{(i)}$, for $i = 1, \ldots, n$, the matrix $R$ with the known parameters, and the probability density function of $e(k)$ (denoted by $p(e)$) are known a priori. Let $\lambda$ be a column vector with the diagonal elements of $\Lambda$. In the identification procedure we will derive estimates $\hat{\theta}$ and $\hat{\lambda}$ of $\theta$ and $\lambda$, respectively.

## 3 Identification of stochastic MPL systems

Suppose that for a given MPL DES of the form (6) we have an input-state sequence $\{(u(k), x(k)\}_{k=1}^{N}$, and that we want to identify the system parameters $\hat{\theta}$ and $\hat{\lambda}$ from this sequence. We make the standard assumption that the input-state sequence is sufficiently rich to capture all the relevant information about the system (see also [17]). We consider the following identification problem:

$$\theta^*, \lambda^* = \arg\min_{\hat{\theta}, \hat{\lambda}} J(\hat{\theta}, \hat{\lambda}) \qquad (8)$$

$$\text{subject to}$$
$$\hat{\lambda} > 0 \tag{9}$$

where

$$J(\hat{\theta},\hat{\lambda}) = \sum_{k=1}^{N-1}\sum_{i=1}^{n} |\mathbb{E}x_i(k+1|k) - x_i(k+1)|^2 \tag{10}$$

In (10), $\mathbb{E}$ denotes the expectation and $\mathbb{E}x_i(k+1|k)$ is the one-step ahead prediction of $x_i$ for event $k+1$, using the knowledge from event $k$.

The first step in the identification procedure is to take a closer look at the one-step-ahead prediction. From (6) and (7) we derive the one-step-ahead prediction

$$\mathbb{E}x_i(k+1|k) \;=\; \mathbb{E}\left(R_{i,\cdot} + \hat{\theta}^T T^{(i)} + e(k)^T \hat{\Lambda} S^{(i)}\right) \otimes p(k)$$

The one-step-ahead prediction error is given by

$$\begin{aligned}
\hat{\eta}_i(k+1,\hat{\theta},\hat{\lambda}) &= \mathbb{E}x_i(k+1|k) - x_i(k+1) \\
&= \mathbb{E}\left(R_{i,\cdot} + \hat{\theta}^T T^{(i)} + e(k)^T \hat{\Lambda} S^{(i)}\right) \otimes p(k) - x_i(k+1) \\
&= \mathbb{E}\max_{j=1,\ldots,m}\Big(R_{i,j} + \hat{\theta}^T T^{(i)}_{\cdot,j} + e(k)^T \hat{\Lambda} S^{(i)}\cdot, j + \\
&\qquad + p_j(k) - x_i(k+1)\Big),
\end{aligned} \tag{11}$$

Now define for a specific realization of the noise vector $e(k)$, the signal

$$\begin{aligned}
\eta_i(k+1,\hat{\theta},\hat{\lambda},e(k)) = \max_{j=1,\ldots,m}\Big(&R_{i,j} + \hat{\theta}^T T^{(i)}_{\cdot,j} + \\
&+ e(k)^T \hat{\Lambda} S^{(i)}_{\cdot,j} + p_j(k) - x_i(k+1)\Big),
\end{aligned}$$

then

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

We will now consider the computation of the $\hat{\eta}$, and derive a subgradient of $\hat{\eta}$ with respect to $\hat{\theta}$ and $\hat{\lambda}$. First we introduce some notation for an easier derivation. Let $\alpha_{ij}(k) = R_{i,j} + p_j(k) - x_i(k+1)$, $\beta_{ij} = T^{(i)}_{\cdot,j}$ and $\Gamma_{ij} = diag(S^{(i)}_{\cdot,j})$, then $[S^{(i)}_{\cdot,j}]^T \cdot \hat{\Lambda} = \hat{\lambda}^T \Gamma_{ij}$ and $\eta(k+1|k)$ can be written as

$$\eta_i(k+1,\hat{\theta},\hat{\lambda},e(k)) \;= \max_{j=1,\ldots,m}\left(\alpha_{ij}(k) + \beta_{ij}^T\hat{\theta} + \hat{\lambda}^T\Gamma_{ij}e(k)\right)$$

Define for any $i$ the sets $\Phi_{ij}(\hat{\theta},\hat{\lambda},k)$, $j = 1,\ldots,m$ such that for all $e \in \Phi_{ij}(\hat{\theta},\hat{\lambda},k)$ there holds:

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

and for any $i$ there holds

$$\bigcup_{j=1}^{m} \Phi_{ij}(\hat{\theta},\hat{\lambda},k) = \mathbb{R}^{n_e}$$

Then

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

$$\begin{aligned}
&= \int_{-\infty}^{\infty}\ldots\int_{-\infty}^{\infty} \eta(k+1,\hat{\theta},\hat{\lambda},e(k))\,p(e)\,de \\
&= \int_{-\infty}^{\infty}\ldots\int_{-\infty}^{\infty} \max_{j=1,\ldots,m}\left(\alpha_{ij}(k) + \beta_{ij}^T\hat{\theta} + \hat{\lambda}^T\Gamma_{ij}e\right)p(e)\,de \\
&= \sum_{j=1}^{m}\int_{e\in\Phi_{ij}(\hat{\theta},\hat{\lambda},k)}\ldots\int \left(\alpha_{ij}(k) + \beta_{ij}^T\hat{\theta} + \hat{\lambda}^T\Gamma_{ij}e\right)p(e)\,de
\end{aligned} \tag{12}$$

where $de = de_1\,de_2\ldots de_{n_e}$.

The following proposition shows that $\hat{\eta}_i(\hat{\theta}(k))$ is convex in $\hat{\theta}$ and $\hat{\lambda}$.

**Proposition 1** *The function $\hat{\eta}_i(k+1,\hat{\theta},\hat{\lambda})$ as defined in (11) is convex in $\hat{\theta}$ and $\hat{\lambda}$, and subgradients $g_{i,\hat{\theta}}(\hat{\theta},\hat{\lambda},k)$ and $g_{i,\hat{\lambda}}(\hat{\theta},\hat{\lambda},k)$ are given by*

$$g_{i,\hat{\theta}}(\hat{\theta},\hat{\lambda},k) \;=\; \sum_{\ell=1}^{m}\left(\int_{e\in\Phi_{i\ell}(\hat{\theta},\hat{\lambda},k)}\ldots\int p(e)\,de\right)\beta_{i\ell}^T \tag{13}$$

$$g_{i,\hat{\lambda}}(\hat{\theta},\hat{\lambda},k) \;=\; \sum_{\ell=1}^{m}\left(\int_{e\in\Phi_{i\ell}(\hat{\theta},\hat{\lambda},k)}\ldots\int e^T p(e)\,de\right)\Gamma_{i\ell}^T \tag{14}$$

**Proof:** Consider vectors $\hat{\theta}_0$ and $\hat{\lambda}_0$ with the same size as $\hat{\theta}$ and $\hat{\lambda}$, respectively. Recall that (cf. (12))

$$\begin{aligned}
\hat{\eta}_i(k+1,\hat{\theta}_0,\hat{\lambda}_0) \;=\; \sum_{\ell=1}^{m}\int_{e\in\Phi_{i\ell}(\hat{\theta}_0,\hat{\lambda}_0,k)}\ldots\int\Big(&\alpha_{i\ell}(k) + \\
&+\beta_{i\ell}^T\hat{\theta}_0 + \hat{\lambda}_0^T\Gamma_{i\ell}e\Big)\,p(e)\,de
\end{aligned}$$

Then, using the fact that $\bigcup\Phi_{i\ell}(\hat{\theta}_0,\hat{\lambda}_0,k) = \mathbb{R}^{n_e}$, there holds for any $\hat{\theta}$ and $\hat{\lambda}$:

$$\begin{aligned}
&\hat{\eta}_i(k+1,\hat{\theta},\hat{\lambda}) = \\
&= \int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty}\max_{j=1,\ldots,m}\left(\alpha_{ij}(k) + \beta_{ij}^T\hat{\theta} + \hat{\lambda}^T\Gamma_{ij}e\right)p(e)\,de \\
&= \sum_{\ell=1}^{m}\int_{e\in\Phi_{i\ell}(\hat{\theta}_0,\hat{\lambda}_0,k)}\ldots\int\max_{j=1,\ldots,m}\Big(\alpha_{ij}(k) + \\
&\qquad\qquad + \beta_{ij}^T\hat{\theta} + \hat{\lambda}^T\Gamma_{ij}e\Big)\,p(e)\,de \\
&\geq \sum_{\ell=1}^{m}\int_{e\in\Phi_{i\ell}(\hat{\theta}_0,\hat{\lambda}_0,k)}\ldots\int\left(\alpha_{i\ell}(k) + \beta_{i\ell}^T\hat{\theta} + \hat{\lambda}\Gamma_{i\ell}e\right)p(e)\,de
\end{aligned} \tag{15}$$

Note that the sets $\Phi_{i\ell}(\cdot,\cdot,k)$ in (15) are computed for $\hat{\theta}_0$ and $\hat{\lambda}_0$, whereas for $\hat{\eta}_i(k+1,\hat{\theta},\hat{\lambda})$, they are computed for $\hat{\theta}$ and $\hat{\lambda}$ (cf. (12)). Now we derive:

$$\begin{aligned}
&\sum_{\ell=1}^{m}\int_{e\in\Phi_{i\ell}(\hat{\theta}_0,\hat{\lambda}_0,k)}\ldots\int\left(\alpha_{i\ell}(k) + \beta_{i\ell}^T\hat{\theta} + \hat{\lambda}^T\Gamma_{i\ell}e\right)p(e)\,de \\
&\quad= \sum_{\ell=1}^{m}\int_{e\in\Phi_{i\ell}(\hat{\theta}_0,\hat{\lambda}_0,k)}\ldots\int\left(\alpha_{i\ell}(k) + \beta_{i\ell}^T\hat{\theta}_0 + \hat{\lambda}_0^T\Gamma_{i\ell}e\right)p(e)\,de \\
&\qquad+ \sum_{\ell=1}^{m}\int_{e\in\Phi_{i\ell}(\hat{\theta}_0,\hat{\lambda}_0,k)}\ldots\int\left(\beta_{i\ell}^T(\hat{\theta}-\hat{\theta}_0)\right)p(e)\,de
\end{aligned}$$

$$
\begin{aligned}
&+\sum_{\ell=1}^{m}\int_{e\in\Phi_{i\ell}(\hat\theta_0,\hat\lambda_0,k)}\cdots\int\Big((\hat\lambda-\hat\lambda_0)^T\Gamma_{i\ell}e\Big)\,p(e)\,de\\
=\;&\sum_{\ell=1}^{m}\int_{e\in\Phi_{i\ell}(\hat\theta_0,\hat\lambda_0,k)}\cdots\int\Big(\alpha_{i\ell}(k)+\beta_{i\ell}^T\hat\theta_0+\hat\lambda_0^T\Gamma_{i\ell}^Te\Big)\,p(e)\,de\\
&+\sum_{\ell=1}^{m}\left(\int_{e\in\Phi_{i\ell}(\hat\theta_0,\hat\lambda_0,k)}\cdots\int p(e)\,de\right)\beta_{i\ell}^T\Big(\hat\theta-\hat\theta_0\Big)\\
&+\sum_{\ell=1}^{m}\left(\int_{e\in\Phi_{i\ell}(\hat\theta_0,\hat\lambda_0,k)}\cdots\int e^Tp(e)\,de\right)\Gamma_{i\ell}^T\Big(\hat\lambda-\hat\lambda_0\Big)\\
=\;&\hat\eta_i(k{+}1,\hat\theta_0,\hat\lambda_0)+g_{i,\hat\theta}(\hat\theta_0,\hat\lambda_0,k)\Big(\hat\theta-\hat\theta_0\Big)\\
&+g_{i,\hat\lambda}(\hat\theta_0,\hat\lambda_0,k)\Big(\hat\lambda-\hat\lambda_0\Big)
\end{aligned}
$$

and we conclude:

$$
\begin{aligned}
&\hat\eta_i(k{+}1,\hat\theta,\hat\lambda)\ge\hat\eta_i(k{+}1,\hat\theta_0,\hat\lambda_0)\\
&+g_{i,\hat\theta}(\hat\theta_0,\hat\lambda_0,k)\Big(\hat\theta-\hat\theta_0\Big)+g_{i,\hat\lambda}(\hat\theta_0,\hat\lambda_0,k)\Big(\hat\lambda-\hat\lambda_0\Big)\qquad(16)
\end{aligned}
$$

From [15] it follows that equation (16) proves that $\hat\eta$ is convex in $\hat\theta$ and $\hat\lambda$ and that $g_{i,\hat\theta}$ and $g_{i,\hat\lambda}$, defined by (13)-(14), are subgradients of $\hat\eta_i$. ■

Now $J(\hat\theta,\hat\lambda)$ in identification problem (8) can be rewritten as

$$
J(\hat\theta,\hat\lambda)=\sum_{k=1}^{N}\sum_{i=1}^{n}\Big|\hat\eta_i(k{+}1,\hat\theta,\hat\lambda)\Big|^2\qquad(17)
$$

The gradients of $J$ become[3]:

$$
\begin{aligned}
\nabla_{\hat\theta}J(\hat\theta,\hat\lambda)&=\sum_{k=1}^{N}\sum_{i=1}^{n}\frac{1}{2}\hat\eta_i(k{+}1,\hat\theta,\hat\lambda)\,g_{i,\hat\theta}(\hat\theta,\hat\lambda,k)\\
\nabla_{\hat\lambda}J(\hat\theta,\hat\lambda)&=\sum_{k=1}^{N}\sum_{i=1}^{n}\frac{1}{2}\hat\eta_i(k{+}1,\hat\theta,\hat\lambda)\,g_{i,\hat\lambda}(\hat\theta,\hat\lambda,k)
\end{aligned}
$$

Note that $J(\hat\theta,\hat\lambda)$ is not convex in $\hat\theta$ and $\hat\lambda$ any more. However we have explicit expression for the gradients, and so gradient search methods with multiple starting points can be used to find the optimal $\hat\theta$ and $\hat\lambda$.

The extension to multi-step ahead prediction is straightforward. Analogous to [20] we can show that a $N_{\rm p}$-step ahead prediction $\hat\eta_i(k{+}j|k,\hat\theta,\hat\lambda)$, $j=1,\ldots,N_{\rm p}$ will be convex in $\hat\theta$ and $\hat\lambda$. The main disadvantage of multi-step ahead predictions is that the stochastic complexity grows dramatically. An elegant way to solve this, is by using variability expansion techniques ([9]) to reduce the computational load (cf. [21]).

## 4 Piecewise polynomial probability density functions

So far, we did not make any assumption on the characterization of probability function $p(e)$. For the computation of the cost criterion we need the values of $\hat\eta(k{+}1,\hat\theta,\hat\lambda)$. If we choose e.g. a Gaussian distribution, they can be calculated from (12) using numerical integration. Numerical integration is usually time-consuming and cumbersome, but can be avoided by choosing polynomial probability density functions (possibly as an approximation of the real probability density function).

Let $p(e)$ be piecewise polynomial functions, so consider sets $P_\ell$ , $\ell=1,\ldots,n_p$, such that for $e\in P_\ell$ the probability density function is given by $p_\ell(e)$, where

$$
p_\ell(e)=\sum_{i_1=0}^{M_1}\sum_{i_2=0}^{M_2}\cdots\sum_{i_m=0}^{M_m}\zeta_{(i_1,i_2,\ldots,i_m)}e_1^{i_1}e_2^{i_2}\cdots e_m^{i_m}
$$

Consider a signal $\eta(k+1,\hat\theta,\hat\lambda)$. Let $E_{ij\ell}(\hat\theta,\hat\lambda,k)=\Phi_{ij}(\hat\theta,\hat\lambda,k)\cap P_\ell$ for $j=1,\ldots,m$, $\ell=1,\ldots,n_p$, then $\hat\eta_i(k{+}1,\hat\theta,\hat\lambda)$ is given by

$$
\begin{aligned}
\hat\eta_i(k{+}1,\hat\theta,\hat\lambda)\;=\;&\sum_{\ell=1}^{n_p}\sum_{j=1}^{m}\int_{e\in E_{ij\ell}(\hat\theta,\hat\lambda,k)}\cdots\int\Big(\alpha_{ij}+\\
&+\beta_{ij}^T\hat\theta+\hat\lambda^T\Gamma_{ij}e\Big)\,p_\ell(e)\,de
\end{aligned}
$$

This is an integral of a polynomial function in $e$ and can be solved analytically for all regions $E_{ij\ell}$ [3, 11].

If piecewise polynomial probability density functions are used as an approximation of "true" non-polynomial probability functions, the quality of the approximation can be improved by increasing the number of sets $n_p$.

## 5 Example

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

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

$$
B(k)=\begin{bmatrix}\theta_3(k)\\ 0\end{bmatrix}
$$

where the parameter vector $\theta$ is given by

$$
\theta\;=\;\begin{bmatrix}\theta_1&\theta_2&\theta_3\end{bmatrix}^T=\begin{bmatrix}0.1&1&0.5\end{bmatrix}^T
$$

These parameters are perturbed by uniformly distributed noise $e_i(k)$ with $-1\le e_i(k)\le 1$, for $i=1,\ldots,5$, and with amplitudes

$$
\lambda\;=\;\begin{bmatrix}\lambda_1&\lambda_2&\lambda_3\end{bmatrix}^T=\begin{bmatrix}0.3&0.3&0.3\end{bmatrix}^T.
$$

In this simulation study we simulate the system for $k=1,\ldots,400$. A parameter estimation is done with the input-state data where the input signal is a staircase signal with an average slope of 0.5, given by

$$
u(k)=1.5\cdot\Big(1+\lfloor k/3\rfloor\Big)
$$

[3] Note that strictly speaking, the chain-rule cannot be applied for subgradients. However, under mild conditions the subgradients $g_{i,\hat\theta}$ and $g_{i,\hat\lambda}$ will be gradients as well.

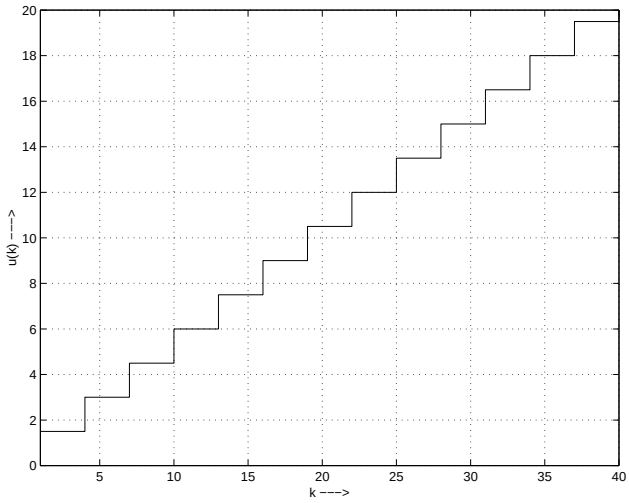


**Figure 1:** Input signal $u(k)$

In Figure 1 input signal $u(k)$ is given for $k = 1,\ldots,20$. As a starting value for our optimization, we use an parameter estimation, based on the technique, described by [1, 5, 13] (in which we assume the noiseless case), adn we obtain:

$$\hat{\theta}_{init} = \begin{bmatrix} -0.1971 & 0.6776 & 0.2047 \end{bmatrix}^T$$

For the noise we use as a starting value:

$$\hat{\lambda}_{init} = \begin{bmatrix} 0.1 & 0.1 & 0.1 \end{bmatrix}^T.$$

Criterion (17), based on the one-step ahead prediction, is minimized using ... We obtain the estimated parameter vector

$$\hat{\theta} = \begin{bmatrix} 0.0781 & 0.9317 & 0.4767 \end{bmatrix}^T$$

and estimated noise amplitude vector

$$\hat{\lambda} = \begin{bmatrix} 0.3591 & 0.5150 & 0.6899 \end{bmatrix}^T$$

## 6 Discussion

In this paper we have derived a system identification procedure for stochastic max-plus-linear systems from observed data. The parameter estimation can be done using a gradient search optimization algorithm. The method works for both structured and fully parameterized state space models. An simulation example has shown that the algorithm makes a good estimate of the system's parameters.

Topics for future research include: development of algorithms for max-plus-linear state space identification, entirely based on input-output data or with only partial state information, and development of methods to obtain good estimates for the system structure and system order.

**Acknowledgements**
Research partially funded by the Dutch Technology Foundation STW project "Model predictive control for hybrid systems" (DMR.5675) and by the European IST project "Modelling, Simulation and Control of Nonsmooth Dynamical Systems (SICONOS)" (IST-2001-37172).

## References

[1] F. Baccelli, G. Cohen, G.J. Olsder, and J.P. Quadrat, *Synchronization and Linearity*. New York: John Wiley & Sons, 1992.

[2] J.L. Boimond, L. Hardouin, and P. Chiron, "A modeling method of SISO discrete-event systems in max-algebra," In *Proceedings of the 3rd European Control Conference*, Rome, Italy, pp. 2023–2026, Sept. 1995.

[3] B. Bueler and A. Enge and K. Fukuda, "Exact Volume Computation for Polytopes: A Practical Study, " In *Proceedings of the conference 'Computational Geometry'*, Münster (Germany), March 1996.

[4] C.G. Cassandras and S. Lafortune, *Introduction to Discrete Event Systems*. Kluwer Academic Publishers, Boston, 1999.

[5] R.A. Cuninghame-Green, *Minimax Algebra*, vol. 166 of *Lecture Notes in Economics and Mathematical Systems*. Berlin, Germany: Springer-Verlag, 1979.

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

[7] B. De Schutter, T.J.J. van den Boom, and V. Verdult, "State space identification of max-plus-linear discrete event systems from input-output data," In *Proceedings of the 2002 Conference on Decision and Control*, Las Vegas, November 1-4, pp. 4024–4029, 2002.

[8] F. Gallot, J. L. Boimond and L. Hardouin, "Identification of simple elements in max–algebra: application to SISO discrete event systems modelisation," In *Proceedings of the 4th European Control Conference*, Brussels, Belgium, 1997.

[9] B. Heidergott, "Variablity Expansion for Performance Characteristic of (Max,Plus)-Linear Systems," In *proceedings of the workshop on Discrete Event Systems (WODES)*, Zaragoza, Spain, 2-4 October, 2002.

[10] Y.C. Ho, *Discrete Event Dynamic Systems: Analyzing Complexity and Performance in the Modern World*. IEEE Press, Piscataway, New Jersey, 1992.

[11] J.B. Lasserre, "Integration on a convex polytope," In *Proceedings of the 1998 American Mathematical Society*, vol.126, no. 8 pp. 2433–2441, August 1998.

[12] L. Ljung. *System Identification : Theory for the User*. Prentice Hall, Upper Saddle River, NJ, 2nd edition, 1999.

[13] Menguy, E., J. L. Boimond, L. Hardouin and J. L. Ferrier. "A First Step Towards Adaptive Control for Linear Systems in Max Algebra," *Discrete Event Dynamic Systems* **10**(4), 347–368, 2000.

[14] J.L. Peterson, “Petri Net Theory and the Modeling of Systems”, Prentice-Hall, Englewood Cliffs, New Jersey, 1981.

[15] R.T. Rockafellar, “Convex Analysis”, Princeton University Press, Princeton, New Jersey, 1970.

[16] G. Schullerus and V. Krebs, “Diagnosis of batch processes based on parameter estimation of discrete event models,” In *Proceedings of the 6th European Control Conference*, Porto, Portugal, pp. 1612–1617, Sept. 2001.

[17] G. Schullerus and V. Krebs, “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, 2001.

[18] G. Schullerus and V. Krebs, “A Method for Estimating the Holding Times in Timed Event Graphs.” In *Proceedings of the 6th Workshop on Discrete Event Systems (WODES’02)*, Zaragossa, 2002.

[19] T.J.J. van den Boom and B. De Schutter, “MPC for perturbed max-plus-linear systems,” *Systems and Control Letters*, 45(1):21–33, January 2002.

[20] T.J.J. van den Boom and B. De Schutter. “MPC for perturbed max-plus-linear systems: A stochastic approach,” In *Proceedings of the 2001 Conference on Decision and Control*, Orlando, FL, USA, pp. 4535–4540, 2001.

[21] T.J.J. van den Boom, B. De Schutter, and B. Heidergott, “Complexity reduction in MPC for stochastic max-plus-linear systems by variability expansion,” In *Proceedings of the 2002 Conference on Decision and Control*, Las Vegas, November 1-4, pp. 3567–3572, 2002.

[22] S. Young and V.K. Garg, “Model uncertainty in discrete event systems,” *SIAM Journal on Control and Optimization*, vol. 33, no. 1, pp. 208–226, Jan. 1995.