Delft University of Technology

Delft Center for Systems and Control

Technical report 15-034

An approximation method for computing the expected value of max-affine expressions*

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

If you want to cite this report, please use the following reference instead:
S.S. Farahani, T. van den Boom, H. van der Weide, and B. De Schutter, "An approximation method for computing the expected value of max-affine expressions," *European Journal of Control*, vol. 27, pp. 17–27, 2016. doi:10.1016/j.ejcon.2015.10.005

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.bartdeschutter.org/abs/15_034.html

An Approximation Method for Computing the Expected Value of Max-Affine Expressions☆

Samira S. Farahani[a,*], Ton van den Boom[a], Hans van der Weide[b], Bart De Schutter[a]

[a] *Delft Center for Systems and Control, Delft University of Technology, Mekelweg 2, 2628 CD Delft, The Netherlands*
[b] *Delft Institute of Applied Mathematics, Delft University of Technology, Mekelweg 4, 2628 CD Delft, The Netherlands*

Abstract

Expected values of max-affine expressions appear in optimization problems for various stochastic systems, such as in model predictive control (MPC) for stochastic max-plus-linear systems, in identification of stochastic max-plus-linear systems, and in control of stochastic monotonic piecewise affine systems. Solving these optimization problems involves the computation of the expected value of the maximum of affine expressions, which will then appear in the objective function or in the constraints. The computation of this expected value can be highly complex and expensive, which also results in a high computation time to solve the optimization problem. Therefore, the focus of this paper is on decreasing the computational complexity of the calculation of these expected values. To this end, we use an approximation method based on the moments of a random variable. We illustrate in an example that this method results in a much lower computation time and a much lower computational complexity than the existing computational methods while still guaranteeing a performance that is comparable to the performance of those methods.

Keywords: Expected value of maximum of affine expressions; discrete-event systems; model predictive control; max-plus linear systems; stochastic disturbance; moments.

1. Introduction

This paper focuses on finding an approximation method to compute the expected value of max-affine expressions. These expected values are often present in the objective functions or constraint functions of the optimization problems for some classes of discrete-event and hybrid systems[1], such as model predictive control (MPC) of stochastic max-plus-linear systems, iden-

☆Research partially funded by the Dutch Technology Foundation STW project "Model-predictive railway traffic management" (11025), and by the European 7th Framework Network of Excellence project "Highly-complex and networked systems (HYCON2)".

*Corresponding author S. S. Farahani. Tel. +31-15-2786891.

Email addresses: s.safaeifarahani@tudelft.nl (Samira S. Farahani), a.j.j.vandenboom@tudelft.nl (Ton van den Boom), j.a.m.vanderweide@tudelft.nl (Hans van der Weide), b.deschutter@tudelft.nl (Bart De Schutter)

[1] In discrete-event systems, the state evolution depends entirely on the occurrence of discrete events over time. Hybrid systems are characterized by the interaction of time-continuous models on the one hand, and logic rules and discrete-event models on the other hand.

tification of stochastic max-plus-linear systems, and control of stochastic monotonic piecewise affine systems. Next, we briefly explain each of these problems in more detail.

Max-plus-linear systems are a special class of discrete-event systems. In broad terms, this class contains discrete-event systems with synchronization but no choice[2] [3]. Models of such systems are based on two fundamental operations, maximization and addition, and they can be applied to both deterministic and stochastic discrete-event systems [3, 14, 25, 30, 33, 44]. In stochastic max-plus-linear systems, modeling errors, noise, and/or disturbances are present; for example, processing times and/or transportation times are then assumed to be stochastic quantities. In practice such stochastic fluctuations can, e.g., be caused by machine failure or depreciation [42]. To control stochastic max-plus-linear systems, one of the efficient and computationally tractable control approaches[3] is model predictive control (MPC) [10, 21, 32]. MPC is an online model-based control approach that relies on a dynamic model of the process and that is capable of handling constraints on both inputs or outputs in a systematic way. Due to its attractive features, MPC has been adopted to discrete-event systems and in particular to max-plus-linear systems [17, 40, 51]. In the optimization problem that appears in MPC for stochastic max-plus-linear systems, the objective function and possibly also one or more constraint functions consist of an expected value of maximum of affine expressions, which imposes a significant computational burden.

Another case in which the expected value of max-affine expressions appears is in the identification of the model parameters of a stochastic max-plus-linear system defined by a state space model. Most identification methods for max-plus-linear discrete-event systems use a transfer function approach [7, 20] while 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 models for identification of *deterministic* max-plus-linear systems, using either the system's Markov parameters or minimizing a prediction error based on input-output data and residuation methods, are presented in [16, 18, 45, 46, 47]. Since in a stochastic max-plus-linear system, the noise and disturbances result in a perturbation of system parameters, in the identification method, the stochastic properties of the systems have to be taken into account. This consideration results in having an expected value of max-affine expressions in the objective function of the identification problem,

The other case in which the expected value of max-affine expressions plays a role is in control of stochastic monotonic piecewise affine (PWA) systems. In PWA systems, the input-state space is partitioned into a collection of non-overlapping polyhedral regions and that in each region, the system dynamics are defined by affine expressions [11, 27, 29, 54, 50]. The class of PWA systems is an important class of hybrid systems since it forms the "simplest" extension of linear systems that can still model non-linear and non-smooth processes with arbitrary accuracy while also being capable to handle hybrid phenomena. Moreover, these systems are equivalent to other classes of hybrid systems such as,mixed logical dynamical systems, max-min-plus-scaling systems, and linear complementarity system [23]. In the control problem of stochastic monotonic PWA systems (MPC is one of the control approaches for such systems [4, 26]), the expected value of max-affine expressions appears in the objective function. Hence, to solve the control

[2]Synchronization requires the availability of several resources or users at the same time, whereas choice appears, e.g., when at a certain time, some user must choose among several resources.

[3]Some related work on other control methods for max-plus-linear systems, such as residuation-based or robust/H_∞ control with the max-plus approach, can be found in [2, 6, 13, 35, 38, 49] and the references therein.

optimization problem, one needs to deal with the computational complexity imposed by this expected value.

One solution approach to compute the expected value of max-affine expressions is numerical integration, which is in general complex and time-consuming. An alternative approach is analytic integration based on piecewise affine probability density functions, which has been proposed in [52]. Although the method of [52] results in an analytic solution, its complexity increases drastically as the number of stochastic variables or the order of the system increase. In [53] an effort is made to reduce the computational complexity by introducing an approximation method, namely the variability expansion that is based on Taylor series. Since this method is analytic and does not resort to simulation, it is in principle possible to compute higher-order moments of performance characteristics of stochastic systems. However, the level of the complexity of the main problem still remains too high.

Accordingly, the aim of this paper is to propose an approximation method to compute the expected value of max-affine expressions with the focus on reducing the computational complexity and the computation time. Here, we propose to approximate the expected value of maximum of affine expressions by its upper bound using higher-order moments of random variables. Since we can compute these moments analytically, this method simplifies the computations considerably. Moreover, due to the special structure of the moments, if this method is applied to a convex optimization setting, it results in a convex optimization problem that can be solved efficiently.

This paper is organized as follows. Section 2 introduces the new approach based on moments of a random variable and describes how it reduces the complexity of computing the expected value of max-affine expressions. In Section 3, the error and the convexity of the approximation method are discussed. Section 4 presents a worked example in which the performance of the approximation method is compared with the one using the analytic integration approach, for MPC for a max-plus-linear system. Finally, Section 5 concludes the paper.

2. The approximation method

Let $\mathcal{S}_{\text{ma}}$ denote the set of max-affine (or max-plus-scaling) functions, i.e., functions f of the form

$$f(z) = \max_{i=1,\dots,n} (\tau_{i,1} z_1 + \cdots + \tau_{i,n_z} z_{n_z} + \xi_i)$$

with $z \in \mathbb{R}^{n_z}$ where n_z denotes the size of the vector z and with constant coefficients $\tau_{i,j}, \xi_i \in \mathbb{R}$. As shown in [52], $\mathcal{S}_{\text{ma}}$ is closed under the operations maximization, addition, and multiplication by a scalar.

Let e denote the vector of independent random variables with a given probability distribution and let n_e denote the number of elements in e. Our aim in this paper is to propose an approximation method to efficiently compute expressions of the form

$$\mathbb{E}[\max_{j=1,\dots,n} (\beta_j + \gamma_j^T e)] \tag{1}$$

where $\beta_j \in \mathbb{R}$ and $\gamma_j \in \mathbb{R}^{n_e}$. This approximation approach is inspired by the relation between the ∞-norm and the p-norm of a vector.

Definition 1 ([22]). *For a vector $x \in \mathbb{R}^n$ and for $p \geq 1$, the p-norm and the ∞-norm of x are defined as:*

$$\begin{aligned} \|x\|_p &= \left(|x_1|^p + \cdots + |x_n|^p\right)^{1/p} \\ \|x\|_\infty &= \max(|x_1|, \dots, |x_n|), \end{aligned} \tag{2}$$

respectively.

The relation between these norms is as follows [22]:

$$\|x\|_\infty \leq \|x\|_p \leq n^{1/p}\|x\|_\infty \tag{3}$$

Moreover, due to the monotonic and linear properties of the expected value, we have

$$\mathbb{E}[\|x\|_\infty] \leq \mathbb{E}[(|x_1|^p + \cdots + |x_n|^p)^{1/p}]. \tag{4}$$

Now, the following proposition shows how we can apply p-norms to find an upper bound for $\mathbb{E}[\max_{j=1,\ldots,n}(\beta_j + \gamma_j^T e)]$.

Proposition 2. *Consider random variables y_j for $j = 1, \ldots, n$ and let $p > 1$. Then*

$$\begin{aligned}\mathbb{E}[\max(y_1, \ldots, y_n)] &\overset{(i)}{\leq} \mathbb{E}[\max(|y_1|, \ldots, |y_n|)] \\ &\overset{(ii)}{\leq} \mathbb{E}[(|y_1|^p + \cdots + |y_n|^p)^{1/p}] \\ &\overset{(iii)}{\leq} \Big(\sum_{j=1}^{n} \mathbb{E}[|y_j|^p]\Big)^{1/p}\end{aligned} \tag{5}$$

Proof. Inequality (i) is straightforward. Inequality (ii) results from (4). Inequality (iii) results from Jensen's inequality for concave functions [8]. Note that we can apply Jensen's Inequality since $\varphi(v) = v^{1/p}$ is a concave function for $p > 1$ and $v > 0$, and in our case the argument v is $\sum_{i=1}^n |y_i|^p$ which is positive. □

Inequality (i) reduces to an equality if all variables x_j are nonnegative. Hence, in order to reduce the error in Inequality (i) for $j = 1, \ldots, n$, we define an offset L such that $x_j = y_j - L$ is almost always positive. Note that if y_j is drawn from a distribution with a finite domain (such as the uniform distribution), L can be defined such that $L \leq y_j$ for $j = 1, \ldots, n$ and hence, Inequality (i) turns into an equality. However, if y_j is drawn from a distribution with an infinite or a left semi-infinite domain[4] (such as the normal distribution), Inequality (i) never reduces to an equality and we can only decrease the error by defining L such that it is less than or equal "almost" all y_j for $j = 1, \ldots, n$. For example if y_j, $j = 1, \ldots, n$ are normally distributed with mean μ_j and variance σ_j^2, then L can be defined as $L = \min_{j=1,\ldots,n}(\mu_j - 3\sigma_j)$. This choice of L is made based on the 3σ-rule, which states that 99.7% of the observations of a normally distributed random variable with mean μ and variance σ^2 fall within the interval $[\mu - 3\sigma, \mu + 3\sigma]$.

Remark 3. *Note that if the number of random variables increases, i.e., for large n, the probability that "almost" all the random variables are larger than or equal to L becomes smaller. This can be explained as follows: for the sake of simplicity assume that the variables y_j, $j = 1, \ldots, n$, are identically distributed. Let p_{prob} denotes the probability that one particular y_j is larger than L. For the normal distribution with $L = \mu_j - 3\sigma_j$, we would then have $p_{prob} \approx 1-(1-0.9973)/2 \approx 0.9987$ since the probability that $y_j \in [\mu_j - 3\sigma_j, \mu_j + 3\sigma_j] \approx 0.9973$ and we only need to consider the interval $[\mu_j - 3\sigma_j, \infty)$. In our derivation, we need that "almost" all y_j are larger than or equal to L. This probability will be p_{prob}^n and hence, if n is large, this probability becomes smaller; for example for $n = 100$, we have $p_{prob}^n = 0.8736$ and for $n = 1000$, we have $p_{prob}^n = 0.2590$. Therefore, if we have a large number of random variables, we should make L smaller and not use the 3σ bound but the 5σ, 7σ, or 9σ bound, depending on the desired accuracy.*

[4] By left semi-infinite domain, we refer to the interval of the form $(-\infty, a)$ for some $a \in \mathbb{R}$.

Remark 4. *In the case that all $y_j,\ j = 1, \ldots, n$ are nonnegative, using the offset L is still useful. Indeed, in that case, L is not needed for Inequality (i) as it will be an equality. However, for Inequality (ii), using an offset L will still reduce the error, which can be illustrated by the following example: Let $y_1 = 1000,\ y_2 = 1001$, then $\max(y_1, y_2) = 1001$. Now for $p = 2$, if $L = 0$ then $(y_1^p + y_2^p)^{1/p} = 1000\sqrt{2}$; however, for $L = 1000$, $((y_1 - L)^p + (y_2 - L)^p)^{1/p} + L = 1000$, which is much closer to $\max(y_1, y_2)$.*

Accordingly, we can rewrite (5) as follows:

$$\begin{aligned}
\mathbb{E}[\max(y_1, \ldots, y_n)] &= \mathbb{E}[\max(y_1 - L, \ldots, y_n - L)] + L \\
&= \mathbb{E}[\max(x_1, \ldots, x_n)] + L \\
&\leq \mathbb{E}[\max(|x_1|, \ldots, |x_n|)] + L \\
&\leq \mathbb{E}[(|x_1|^p + \cdots + |x_n|^p)^{1/p}] + L \\
&\leq \Big(\sum_{j=1}^{n} \mathbb{E}[|x_j|^p] \Big)^{1/p} + L. \qquad (6)
\end{aligned}$$

Remark 5. *For a positive even integer $p = 2q,\ q \in \mathbb{N} \setminus \{0\}$, we have $\mathbb{E}[x^p] = \mathbb{E}[|x|^p]$. Hence, without loss of generality, we can use $\mathbb{E}[x^p]$ in (6). So from now on, p is an even integer larger than or equal to 2.*

Therefore, considering Inequality (6) and Remark 5, we can approximate $\mathbb{E}[\max_{j=1,\ldots,n}(\beta_j + \gamma_j^T e)]$ in (1) by an upper bound. Let $y_j = \beta_j + \gamma_j^T e$; hence, an upper bound can be defined as follows:

$$\mathfrak{U}\Big(\mathbb{E}[\max(y_1, \ldots, y_n)]\Big) = \Big(\sum_{j=1}^{n} \mathbb{E}[(\beta_j + \gamma_j^T e - L)^p] \Big)^{1/p} + L \qquad (7)$$

for p a positive even integer and for independent random variables $y_j,\ j = 1, \ldots, n$.

In the approximation function (7), we have to compute the p-th moment of each random variable $x_j = y_j - L = \beta_j + \gamma_j^T e - L,\ j = 1, \ldots, n$. The p-th moment of a real-valued random variable x is defined as follows:

$$\mathbb{E}[x^p] = \int_{-\infty}^{\infty} x^p f(x) dx \qquad (8)$$

where $f(\cdot)$ is the probability density function of x. In general, moments of a random variable can be finite or infinite. Hence, to be able to usefully apply $\mathfrak{U}\Big(\mathbb{E}[\max_{j=1,\ldots,n}(\beta_j + \gamma_j^T e)]\Big)$ as an approximation of $\mathbb{E}[\max_{j=1,\ldots,n}(\beta_j + \gamma_j^T e)]$, we need to consider random variables (i.e., each $x_j,\ j = 1, \ldots, n$) with finite moments for which a closed-form expression exists. Furthermore, each random variable $x_j,\ j = 1, \ldots, n$ is indeed a sum of the few random variables since $\gamma_j^T e = \gamma_{j,1} e_1 + \cdots + \gamma_{j,n_e} e_{n_e}$. Therefore, in order to determine the distribution of each x_j, we need to either consider random variables with distributions that are preserved under summation such as the normal distribution, the Poisson distribution, and the gamma distribution [5, 43] or random variables that their sum has a known distribution with finite moments for which a closed-form expression exists, such as n i.i.d random variables with uniform distribution $U(0, 1)$ since their sum has the Irwin-Hall distribution [28], which has finite moments with closed-form expressions.

As an example, we assume in this paper that each element of the stochastic vector e, namely $e_t,\ t = 1, \ldots, n_e$ is normally distributed with mean μ_t and variance σ_t^2, i.e., $e_t \sim \mathcal{N}(\mu_t, \sigma_t^2)$.

Remark 6. *For a normally distributed stochastic vector e, the random variable* $x_j = \beta_j + \gamma_j^T e - L$ *is also normally distributed with mean* $\beta_j - L + \gamma_j^T \mu^T$ *and variance* $(\gamma_j^2)^T (\sigma^2)^T$ *where* $\mu = [\mu_1, \ldots, \mu_{n_e}]^T$ *and* $\sigma^2 = [\sigma_1^2, \ldots, \sigma_{n_e}^2]^T$*. This is obtained using the property of the normal distribution that sum of the independent normally distributed random variables has also a normal distribution with a new mean and variance [19].*

Furthermore, according to [55], the p-th moment of a normally distributed scalar random variable x with mean μ and variance σ^2 has a closed form that can be expressed as:

$$\mathbb{E}[x^p] = \sigma^p i^{-p} H_p(i\mu/\sigma) \tag{9}$$

where

$$H_p(x) = p! \sum_{l=0}^{p/2} \frac{(-1)^l x^{p-2l}}{2^l l!(p-2l)!} \tag{10}$$

is the p-th Hermite polynomial obtained from equations (26.2.51) and (22.3.11) in [1] and $p/2 \in \mathbb{N} \setminus \{0\}$ since, by assumption, p is an even integer in our case. Note that the right-hand side of (9) is in fact real because $H_p(x)$ contains only even powers of x if p is even.

As a result, we can then replace $\mathbb{E}[(\beta_j + \gamma_j^T e - L)^p]$ in (7) by the expression in (9) with the appropriately defined μ_j and σ_j^2 related to each random variable $x_j = \beta_j + \gamma_j^T e - L$. Note that by using (9), the computational complexity of this approximation method is of order $O\big(n \cdot n_e \cdot p\big)$ and hence, it increases polynomially as n, n_e, or p increase. The computational complexity of the two other approaches proposed in [52, 53] are also presented in Section 6.2 of the appendix for the sake of comparison.

Remark 7. *A similar approximation method for calculating the expected value of the maximum of a finite set of random variables has been proposed in [12]. In that approach, the random variables are assumed to be normally distributed (possibly dependent) or they are approximated by normally distributed variables. The expected value of the maximum of finite set of these variables can be obtained using an iterative approximation method. The differences between this method and the method proposed in the current paper are: 1) in our method the expected value and hence, the objective function can be computed analytically while the method of [12] they are not computed fully analytic due to the presence of the cumulative distribution functions of the normal distribution in the formulas for which the look-up tables are required; 2) most importantly, the approximation method in [12] does not provide an upper bound for the exited value while our approach does. This is important in the optimization context, since by minimizing the approximate objective function, obtained by using our proposed approximation method, we guarantee that the optimal approximate solution gets closer to the exact optimal solution.*

3. On the error and the convexity of the approximation method

In this section, we show that the error due to the proposed approximation method is bounded from below and from above. Moreover, we prove that the approximation function (7) is a convex function of its variables.

3.1. Error of the approximation method

In this section, we show that the error caused by approximating $\mathbb{E}[\max(x_1,\dots,x_n)]$ by its upper bound $\mathfrak{U}\Big(\mathbb{E}[\max(x_1,\dots,x_n)]\Big) = \big(\sum_{j=1}^{n}\mathbb{E}[(x_j-L)^p]\big)^{1/p}+L$ (cf. (7)) is bounded. Note that $\mathbb{E}[\max(x_1,\dots,x_n)]$ is bounded from above and from below. Indeed, an upper bound has been presented in (7) and a lower bound can be obtained using Jensen's inequality for convex functions, the max function in this case. Let $\mathfrak{L}\Big(\mathbb{E}[\max(x_1,\dots,x_n)]\Big) = \max(\mathbb{E}[x_1],\dots,\mathbb{E}[x_n])$ denote this lower bound. Hence,

$$\mathfrak{L}\Big(\mathbb{E}[\max(x_1,\dots,x_n)]\Big) \le \mathbb{E}[\max(x_1,\dots,x_n)] \le \mathfrak{U}\Big(\mathbb{E}[\max(x_1,\dots,x_n)]\Big) \tag{11}$$

Consequently, the error of approximating $\mathbb{E}[\max(x_1,\dots,x_n)]$ by its upper bound is always bounded as follows:

$$0 \le \mathfrak{U}\Big(\mathbb{E}[\max(x_1,\dots,x_n)]\Big) - \mathbb{E}[\max(x_1,\dots,x_n)] \le \mathfrak{U}\Big(\mathbb{E}[\max(x_1,\dots,x_n)]\Big) - \mathfrak{L}\Big(\mathbb{E}[\max(x_1,\dots,x_n)]\Big) \tag{12}$$

and since in our case $x_j,\ j=1,\dots,n$ are assumed to have finite moments, this upper bound is finite and the error of the approximation cannot be larger than this value.

Alternatively, we can introduce another upper bound for the approximation error as follows. We can split the error introduced by the proposed approximation method into three parts corresponding to the three inequalities in Proposition 2. Since we have defined $x_j = y_j - L$, in the case that $y_j, j=1,\dots,n$ is drawn from a distribution with a finite domain, the error due to (i) becomes zero by choosing $L = \min_{j=1,\dots,n}(y_j)$; moreover, in the case that $y_j, j=1,\dots,n$ is drawn from a distribution with an infinite domain, the error due to (i) approaches 0 if L becomes smaller and smaller. The error due to (ii) approaches 0 if $p \to +\infty$, since by definition $\|x\|_\infty = \lim_{p\to+\infty}\|x\|_p$. The third error, which is in fact the error of Jensen's inequality, needs more discussion. Note that in the case of infinite domain, we will only obtain an approximate upper bound for the error by applying the approach.

In [48, Theorem 3] an upper bound for the absolute error of Jensen's inequality is presented as follows: for a differentiable, concave function f on an interval $[a,b]$ we have

$$0 \le f(\mathbb{E}[x]) - \mathbb{E}[f(x)] \le \max_{q\in[0,1]}[f(qa+(1-q)b) - qf(a) - (1-q)f(b)] =: T_f(a,b)\ .$$

By substituting $f(x) = x^{1/p}$ in the above formula and determining the maximum for $q \in [0,1]$, we obtain

$$T_f(a,b) = \left(\frac{a-b}{p(a^{\frac{1}{p}}-b^{\frac{1}{p}})}\right)^{\frac{1}{p-1}} - \left(\frac{1}{a-b}\left[(a^{\frac{1}{p}}-b^{\frac{1}{p}})\Big(\frac{a-b}{p(a^{\frac{1}{p}}-b^{\frac{1}{p}})}\Big)^{\frac{p}{p-1}} - a^{\frac{1}{p}}b + ab^{\frac{1}{p}}\right]\right) \tag{13}$$

and for different values of a, b and p and by applying l'Hôpital's rule, we obtain the following:

if ($a \to \infty$ or $b \to \infty$) and ($p < \infty$) then $T_f(a,b) \to \infty$
if ($a < \infty$ and $b < \infty$) and ($p \to \infty$) then $T_f(a,b) \to 0$

As mentioned in Section 2, we assume in this paper that random variables $x_1,\dots,x_n$ are independent and normally distributed. Hence, the argument x defined as $\sum_{j=1}^{n} x_j^p$ is not bounded. However, each $x_j = y_j - L$ is in the interval $[\mu_j - 3\sigma_j - L, \mu_j + 3\sigma_j - L]$ with probability 99.7%

(cf. Section 2). Hence, we compute an approximate error[5] by only considering the case that x_j belongs to the interval $[c_{1j}, c_{2j}]$, where $c_{1j} := \mu_j - 3\sigma_j - L$ and $c_{2j} := \mu_j + 3\sigma_j - L$. Since we have $L = \min_j(\mu_j - 3\sigma_j)$, it follows that $0 \leq c_{1j} \leq c_{2j}$ and consequently $c_{1j}^p \leq c_{2j}^p$. Consequently, for $x = \sum_{j=1}^n x_j^p$ we have:

$$\underbrace{\sum_{j=1}^{n} c_{1j}^p}_{a} \leq \underbrace{\sum_{j=1}^{n} x_j^p}_{x} \leq \underbrace{\sum_{j=1}^{n} c_{2j}^p}_{b} \tag{14}$$

Recall that the error caused by Inequality (*ii*) in Proposition 2 approaches 0 as $p \to \infty$. This suggests that in order to get a good approximation, p should be selected very large. However, since in our case both a and b depend on p (cf. (14)), we need a more careful investigation to study the effect of $p \to \infty$ on $T_f(a, b)$. To this end, let $\alpha = \max_{j=1,\ldots,n}(c_{1j})$ and $\beta = \max_{j=1,\ldots,n} c_{2j}$. Since $c_{1j} < c_{2j}$ for all j, we conclude that $\alpha < \beta$. Denote the number of $c_{1j},\ j = 1, \ldots, n$ that are equal to α by A and the number of $c_{2j},\ j = 1, \ldots, n$ that are equal to β by B. Now, for a large p, we can rewrite a and b as $a \approx A\alpha^p$ and $b \approx B\beta^p$. Using this notation, we obtain the following proposition.

Proposition 8. *Considering our assumption that p is a positive even integer and that $a \approx A\alpha^p,\ b \approx B\beta^p$ for a large p with A, B positive integers and $0 \leq \alpha < \beta$, we have the following result:* $\lim_{p\to\infty} T_f(\alpha, \beta) = \beta$

Proof. The proof is straightforward using l'Hôpital's rule. □

Accordingly, considering the effect of the second error (which becomes smaller as $p \to +\infty$) and the third error (which depends on the magnitude of β as $p \to +\infty$), there is a trade-off between the choice of p and the magnitude of the approximation error (see also Section 4). Note that this error bound is only an approximation, since we leave out the cases where $x_j > \mu_j + 3\sigma_j - L$ and $x_j < \mu_j - 3\sigma_j - L$ for $j = 1, \ldots, n$.

3.2. Convexity of the approximation method

In this section, we also prove that the approximation function (cf. (7))

$$\mathfrak{U}\Big(\mathbb{E}[\max_{j=1,\ldots,n}(\beta_j - L + \gamma_j^T e)]\Big) = \Big(\sum_{j=1}^{n} \mathbb{E}[(\beta_j - L + \gamma_j^T e)^p]\Big)^{1/p} + L$$

with $\beta_j = a_j + b_j^T w$, where w denotes a general control or optimization variable, is convex in w. To this end, let $\phi_j(w) = a_j + b_j^T w + \gamma_j^T e - L$, which is an affine and so a convex function in w. Hence, the p-th moment of $\phi_j(w)$, i.e., $\mathbb{E}[(\phi_j(w))^p]$ can be defined as:

$$\mathbb{E}[(\phi_j(w))^p] = \int_{-\infty}^{+\infty} \cdots \int_{-\infty}^{+\infty} (\phi_j(w))^p\ f(e)de$$

where $f(e)$ is the probability density function of e. Note that to the random variable e, we only assign distributions that have finite moments; hence, $\mathbb{E}[(\phi_j(w))^p]$ is finite as well. Consider the following proposition:

[5] In case that the obtained approximate error bound is not small enough, one can improve it by considering 5σ- or 7σ-rule.

Proposition 9. $\left(\mathbb{E}[(\phi_j(w))^p]\right)^{1/p}$, *with* ϕ_j *an affine function in* w, *is a convex function of* w.

Proof. If we show that

$$\begin{aligned}&\left\{\mathbb{E}[(\phi_j(\lambda w_1 + (1-\lambda)w_2))^p]\right\}^{1/p}\\ &\leq \lambda\left\{\mathbb{E}[(\phi_j(w_1))^p]\right\}^{1/p} + (1-\lambda)\left\{\mathbb{E}[(\phi_j(w_2))^p]\right\}^{1/p}\end{aligned}$$

for any two points w_1 and w_2 in the domain of $\mathbb{E}[(\phi_j(w))^p]$ and for any $0 \leq \lambda \leq 1$, then the proof is complete. Since ϕ_j is an affine function in w, we have

$$\phi_j(\lambda w_1 + (1-\lambda)w_2) = \lambda\phi_j(w_1) + (1-\lambda)\phi_j(w_2)$$

Therefore, from the Minkowski inequality for functions [39, Chapter 5] and keeping in mind that p is an even integer we obtain:

$$\begin{aligned}\left(\mathbb{E}[(\phi_j(\lambda w_1 + (1-\lambda)w_2))^p]\right)^{1/p} &= \left(\mathbb{E}[(\lambda\phi_j(w_1) + (1-\lambda)\phi_j(w_2))^p]\right)^{1/p}\\ &\leq \lambda\left(\mathbb{E}[(\phi_j(w_1))^p]\right)^{1/p} + (1-\lambda)\left(\mathbb{E}[(\phi_j(w_2))^p]\right)^{1/p} \qquad (15)\end{aligned}$$

So the inequality holds true and consequently, $\left(\mathbb{E}[(\phi_j(w))^p]\right)^{1/p}$ is a convex function in w. □

This proposition implies that

$$\mathfrak{U}\left(\mathbb{E}[\max_{j=1,\ldots,n}\phi_j(w)]\right) = \left(\sum_{j=1}^{n}\mathbb{E}[(\phi_j(w))^p]\right)^{1/p} + L \qquad (16)$$

is a convex function in w. Its subgradient can be defined as follows:

$$\begin{aligned}\frac{\partial}{\partial w}\mathfrak{U}\left(\mathbb{E}[\max_{j=1,\ldots,n}\phi_j(w)]\right) = &\left(\sum_{j=1}^{n}\mathbb{E}[(a_j + b_j^T w + \gamma_j^T e - L)^p]\right)^{1/p-1}\\ &\cdot\sum_{j=1}^{n} b_j\mathbb{E}[(a_j + b_j^T w + \gamma_j^T e - L)^{p-1}].\end{aligned}$$

The convexity of the approximation function $\mathfrak{U}\left(\mathbb{E}[\max_{j=1,\ldots,n}\phi_j(w)]\right)$ is useful in the case that we have a convex optimization setting. In that case, by approximating the objective function using (7), we obtain again a convex optimization problem that can be solved efficiently using convex optimization algorithms such as interior point methods [41, 56].

4. Example

In this section, we show how the proposed approximation method will improve the computational efficiency of an stochastic optimization problem in which the objective function is defined as the expected value of max-affine expressions. We will apply the approximation function (7) to approximate this objective function in the optimization problem and we will compare the obtained results with the ones of some other computational methods.

We reconsider the example of [17] and we assume that the processing time of machine 2 and 3 are both perturbed by an stochastic vector $e(k)$ at each event step k. Moreover, we assume that the element of $e(k)$ are normally distributed. In this example, we check whether our approximation method works efficiently when it is applied to the MPC[6] optimization problem of stochastic max-plus-linear systems. To compute the objective function, which is an expected value of max-plus expressions, that appears in the MPC optimization problem, we apply different methods namely, numerical integration based on Monte Carlo methods [15], nominal MPC, Monte Carlo simulation, and the approximation method proposed in this paper. Afterward, we compare the performance of the MPC controller using these four methods.

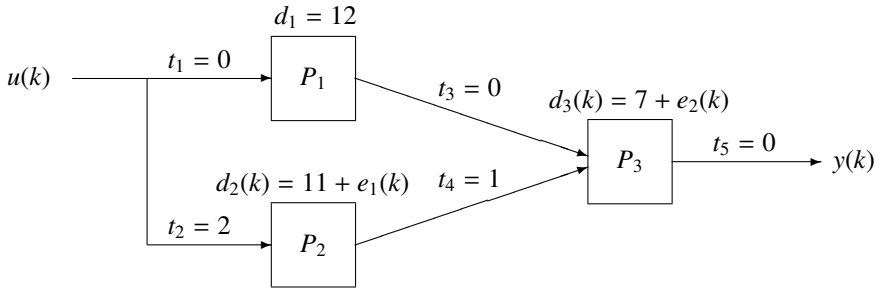

Figure 1: A production system.

Consider the simple manufacturing system of Figure 1. Define:

$u(k)$	:	time instant at which the raw material is fed to the system for the k-th time
$y(k)$	:	time instant at which the k-th product leaves the system
$x_i(k)$	:	time instant at which machine i starts for the k-th time
$t_j(k)$	:	transportation time on link j for the k-th batch.
$d_i(k)$	:	processing time on machine i for the k-th batch.

The system equations are given by

$$\begin{aligned}
x_1(k) &= \max(x_1(k-1) + d_1,\ u(k) + t_1\ ) \\
x_2(k) &= \max(x_2(k-1) + d_2(k-1),\ u(k) + t_2\ ) \\
x_3(k) &= \max(x_1(k) + d_1 + t_3, x_2(k) + d_2(k) + t_4,\ x_3(k-1) + d_3(k-1)\ ) \\
&= \max(x_1(k-1) + 2d_1 + t_3,\ x_2(k-1) + d_2(k-1) + d_2(k) + t_4, \\
&\qquad x_3(k-1) + d_3(k-1), u(k) + d_1 + t_1 + t_3,\ u(k) + d_2(k) + t_2 + t_4\ ) \\
y(k) &= x_3(k) + d_3(k) + t_5
\end{aligned}$$

and in matrix notation this becomes[7]

$$x(k) = A(k) \otimes x(k-1) \oplus B(k) \otimes u(k)$$

[6] A brief description of MPC for max-plus-linear systems has been presented in Section 6.1 of the appendix and for more details the interested reader is referred to [52].

[7] In max-plus notation for A and B matrices, $(A \oplus B)_{ij} = a_{ij} \oplus b_{ij} = \max(a_{ij}, b_{ij})$ and $(A \otimes C)_{ij} = \bigoplus_{k=1}^{n} a_{ik} \otimes c_{kj} = \max_{k=1,\ldots,n}(a_{ik} + c_{kj})$.

$$y(k) \quad = \quad C(k) \otimes x(k) \ .$$

where the system matrices A, B and C are given by

$$A(k) = \begin{bmatrix} d_1 & \varepsilon & \varepsilon \\ \varepsilon & d_2(k-1) & \varepsilon \\ 2d_1 + t_3 & d_2(k-1) + d_2(k) + t_4 & d_3(k-1) \end{bmatrix},$$

$$B(k) = \begin{bmatrix} t_1 \\ t_2 \\ \max(d_1 + t_1 + t_3, d_2(k) + t_2 + t_4) \end{bmatrix}, \ C(k) = \begin{bmatrix} \varepsilon & \varepsilon & d_3(k) \end{bmatrix}.$$

The objective function, defined as

$$J(k) = \sum_{j=0}^{N_\mathrm{p}-1} \sum_{i=1}^{n_y} \mathbb{E}[\max(y_i(k+j) - r_i(k+j), 0)] - \lambda \sum_{j=0}^{N_\mathrm{p}-1} \sum_{l=1}^{n_u} u_l(k+j), \tag{17}$$

will be optimized for the prediction horizon $N_\mathrm{p} = 3$, $\lambda = 0.05$, the reference signal $r(k) = 6 + 20 \cdot k$ where $k = 1, \ldots, 40$, and $x(0) = [0\ 0\ 10]^T$. The signal d_2 and d_3 are assumed to be corrupted by Gaussian noise: $d_2(k+\ell) = 11 + \tilde{e}_1(k+\ell)$ and $d_3(k+\ell) = 7 + \tilde{e}_2(k+\ell)$ where $\tilde{e}_1(k+\ell), \tilde{e}_2(k+\ell)$ are assumed to be i.i.d and $\tilde{e}_1(k+\ell), \tilde{e}_2(k+\ell) \sim \mathcal{N}(0,1)$ for $\ell = -1, \ldots, N_\mathrm{p} - 1$. Hence, the vector $e(k) = [\ d_2(k-1) \quad \ldots \quad d_2(k+N_\mathrm{p}-1) \quad d_3(k-1) \quad \ldots \quad d_3(k+N_\mathrm{p}-1)\ ]^T$ consists of independent stochastic random variables. Now, we have

$$\begin{aligned}
\max(y(k) - r(k), 0) = \max(&\eta_1 + e_6, \eta_2 + e_1 + e_2 + e_6, \eta_3 + e_5 + e_6, \eta_4 + e_6, \\
&\eta_5 + e_2 + e_6, 0) \\
\max(y(k+1) - r(k+1), 0) = \max(&\eta_6 + e_7, \eta_7 + e_7, \eta_8 + e_1 + e_2 + e_3 + e_7, \\
&\eta_9 + e_2 + e_3 + e_7, \eta_{10} + e_6 + e_7, \\
&\eta_{11} + e_1 + e_2 + e_6 + e_7, \eta_{12} + e_5 + e + 6 + e_7, \eta_{13} + e_6 + e_7 \\
&\eta_{14} + e_2 + e_7 + e_7, \eta_{15} + e_7, \eta_{16} + e_3 + e_7, 0) \\
\max(y(k+2) - r(k+2), 0) = \max(&\eta_{17} + e_8, \eta_{18} + e_8, \eta_{19} + e_8, \eta_{20} + e_1 + e_2 + e_3 + e_4 + e_8 \\
&\eta_{21} + e_2 + e_3 + e_4 + e_8, \eta_{22} + e_3 + e_4 + e_8, \\
&\eta_{23} + e_7 + e_8, \eta_{24} + e_7 + e_8, \eta_{25} + e_1 + e_2 + e_3 + e_7 + e_8, \\
&\eta_{26} + e_3 + e_3 + e_7 + e_8, \eta_{27} + e_6 + e_7 + e_8, \\
&\eta_{28} + e_1 + e_2 + e_6 + e_7 + e_8, \eta_{29} + e_5 + e_6 + e_7 + e_8, \\
&\eta_{30} + e_6 + e_7 + e_8, \eta_{31} + e_2 + e_6 + e_7 + e_8, \eta_{32} + e_7 + e_8, \\
&\eta_{33} + e_3 + e_7 + e_8, \eta_{34} + e_8, \eta_{35} + e_4 + e_8, 0)
\end{aligned} \tag{18}$$

where[8] $\eta_1, \ldots, \eta_{35}$ are sums of deterministic values and $e_1, \ldots, e_8$ are the entries of $e(k)$. Note that for independent normally distributed random variables $e_1, \ldots, e_8$ with mean μ_i and variance σ_i^2, $\sum_{i=1}^8 e_i$ is also normally distributed with mean $\mu = \sum_{i=1}^8 \mu_i$ and variance $\sigma^2 = \sum_{i=1}^8 \sigma_i^2$. Hence, for each equation in (18), we can define the scalars L_1, L_2, and L_3 respectively, as follows:

$$L_1 = \min(\eta_1 - 3 \cdot 1, \eta_2 - 3 \cdot \sqrt{3}, \eta_3 - 3 \cdot \sqrt{2}, \eta_4 - 3 \cdot 1, \eta_5 - 3 \cdot \sqrt{2}, 0)$$

[8] We have omitted the argument k for brevity. Note also that the number of affine expressions, i.e., n in (7), is 6,12, and 20, respectively (cf. (18)).

$$
\begin{aligned}
L_2 = \min(&\eta_6 - 3\cdot 1, \eta_7 - 3\cdot 1, \eta_8 - 3\cdot 2, \eta_9 - 3\cdot \sqrt{3}, \eta_{10} - 3\cdot \sqrt{2}, \eta_{11} - 3\cdot 2,\\
&\eta_{12} - 3\cdot \sqrt{3}, \eta_{13} - 3\cdot \sqrt{2}, \eta_{14} - 3\cdot \sqrt{3}, \eta_{15} - 3\cdot 1, \eta_{16} - 3\cdot \sqrt{2}, 0)\\
L_3 = \min(&\eta_{17} - 3\cdot 1, \eta_{18} - 3\cdot 1, \eta_{19} - 3\cdot 1, \eta_{20} - 3\cdot \sqrt{5}, \eta_{21} - 3\cdot 2, \eta_{22} - 3\cdot \sqrt{3}, \eta_{23} - 3\cdot \sqrt{2},\\
&\eta_{24} - 3\cdot \sqrt{2}\eta_{25} - 3\cdot \sqrt{5}, \eta_{26} - 3\cdot 2, \eta_{27} - 3\cdot \sqrt{3}, \eta_{28} - 3\cdot \sqrt{5},\\
&\eta_{29} - 3\cdot 2, \eta_{30} - 3\cdot \sqrt{3}, \eta_{31} - 3\cdot 2, \eta_{32} - 3\cdot \sqrt{2}, \eta_{33} - 3\cdot \sqrt{2},\\
&\eta_{34} - 3\cdot 1, \eta_{35} - 3\cdot \sqrt{2}, 0)
\end{aligned}
$$

Recall that this choice is based on the 3σ-rule (cf. Section 2).

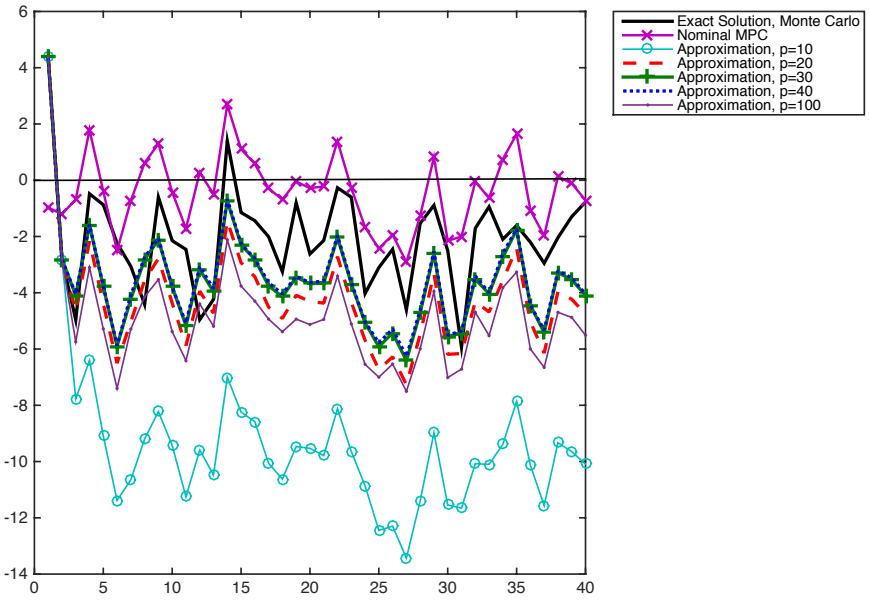

Figure 2: Due date error $y(k) - r(k)$ for the closed-loop system using nominal MPC, numerical integration, and the approximation method to compute the expected value in the objective function (17).

Figure 2 shows the difference between the output signal y and the due date signal r for the closed-loop simulation. The optimization has been done using sequential quadratic programming (SQP) and the *fmincon* optimization function in Matlab with one initial value since the objective functions (both the exact and the approximate one) are convex and hence, the global minimum can be reached from any initial point. In Figure 2, the "Exact solution" is obtained by using numerical integration to compute the expected value in the objective function (17). We have also used Monte Carlo simulation to compute the expected value that appears in the objective function. Note that the plot of the Monte Carlo simulation coincides with the one of the "Exact solution". The "Nominal MPC" solution is obtained by applying deterministic max-plus MPC [17] to the nominal system while computing the optimal input sequence. As a result, the due dates will be violated most of the time in this case and we have late deliveries. This is due to the

fact that we compute the optimal input in the absence of noise and hence we cannot minimize the effect of noise. Finally, the "Approximation" solution is obtained by using the proposed approximation method to compute the expected value in the objective function (17).

We have chosen different values of p to find out which one gives the closest result to the exact solution considering the effect of the error of the approximation method (cf. Section 3.1). As can be seen, or $p = 20,\ 30$, and 40, the result of the closed-loop controlled system using the approximation method are the closest to the one of the "Exact solution". If we further increase p from 50 to 90, the results does not improve and on the contrary, they get closer to the one of $p = 20$ (these plots are not shown here for the clarity of the figure). For $p = 100$, as shown in Figure 2, the difference between the output and reference signal goes even below the one of $p = 20$, and hence, gets further away from the exact solution. This observation shows clearly that a larger p does not always give a better result.

In Table 1, the total optimization time[9] for closed-loop simulation over 40 event steps using numerical integration approach, nominal MPC, Monte Carlo simulation, and the approximation method for different values of p are reported. For each method, MPC has been run in closed loop 20 times, each time with a different noise realization. Note that at each time, the same noise realization has been used for all the methods. In Table 1, we report the mean value of the performance criterion J_{tot} over the entire simulation period, where $J_{\text{tot}} = \sum_{k=1}^{40} (\max(y(k) - r(k), 0) - \lambda u(k))$. In addition, the relative error[10] of the mean value of J_{tot} using numerical integration versus using the other methods is presented.

Methods for computing the objective function	Computation time[s]	Mean value of J_{tot}	Relative error of mean value of J_{tot}
Numerical integration	864.95	−780.2693	–
Nominal MPC	0.28	−735.4663	5.74%
Monte Carlo	1509.01	−780.5834	0.04%
Approximation method:			
$p = 10$	11.39	−767.3861	1.65%
$p = 20$	15.15	−777.7221	0.33%
$p = 30$	20.21	−779.0438	0.16%
$p = 40$	30.63	−779.1575	0.14%
$p = 100$	63.72	−776.5324	0.48%

Table 1: The computation time, the mean value of J_{tot}, and the relative error of the mean value of J_{tot} using different methods to compute the expected value in the objective function.

Here again, compared to numerical integration, nominal MPC results in a very large objective function value. Hence, despite the very short computation time, nominal MPC is not reliable to be applied. The computation time of the optimization procedure using the approximation method is on average about a factor 50 smaller than the one using Monte Carlo simulation and it is about

[9] These times are obtained running MATLAB R2014b on a 2.6 GHz Intel Core i5 processor.

[10] The relative error is defined here as $\frac{|J_{\text{tot,app}} - J_{\text{tot,ni}}|}{|J_{\text{tot,ni}}|}$ where $J_{\text{tot,app}}$ is obtained using the approximation method and $J_{\text{tot,ni}}$ is obtained using the numerical integration approach.

a factor 30 smaller than the one using numerical integration. The number of samples of the Monte Carlo simulation has chosen to be 10^5 since for 10^7 samples, although the relative error of the objective function will be 0.007% compared to 0.04% for 10^5 samples, the computation time will be a factor 100 larger than the one for 10^5 samples. Note that since we use the approximation function (7), increasing the value of p has only a linear effect on the computation time and even for $p = 100$, the computation time is still much lower than the one using Monte Carlo simulation or numerical integration. As a result, by comparing the CPU time of these three methods, we can conclude that the approximation method is considerably faster than numerical integration and Monte Carlo simulation. Moreover, this table shows again that – considering the relative error for different values of p – a larger p does not always make the approximation better. This is due to the fact that although a larger p decreases the error of inequality (ii) in (5), it increases the error of inequality (iii), i.e., the Jensen's inequality (cf. Section 3.1). Consequently, one needs to find the appropriate value of p that gives the best approximation result, which has been done here by means of experiments.

It is also possible to compare the time required to obtain a certain accuracy. For the approximation method, we can improve the accuracy by choosing smaller L, e.g., by selecting 5σ or 7σ; however, the computation time will be still similar to the ones presented in Table 1 for different values of p. For the Monte Carlo approach, it is possible to decrease the accuracy to the level of the one of the approximation method that has the smallest relative error, i.e., for $p = 40$, and then compare the required computation time for obtaining such accuracy. This has been done by fixing the termination time of the SQP iteration. The results are presented in Table 2.

Methods for computing the objective function	Computation time[s]	Mean value of J_{tot}	Relative error of mean value of J_{tot}
Numerical integration	864.95	−780.2693	–
Monte Carlo	107.32	−779.0989	0.15%
Approximation method:$p = 40$	30.63	−779.1575	0.14%

Table 2: Comparing the computation time, the mean value of J_{tot}, and the relative error of the mean value of J_{tot} using Monte Carlo method with a fixed accuracy level, numerical integration and the approximation method with $p = 40$.

As shown here, even if we sacrifice the accuracy of the Monte Carlo method, the computation time is still much higher than the one of the approximation method for different values of p (cf. Table 1).

Figure 3 shows the evolution of $J^*_{\text{app}}(k) - J^*(k)$, where $J^*(k)$ is obtained by solving the optimization problem in closed loop using numerical integration and $J^*_{\text{app}}(k)$ is computed using the approximation method. Note that to compute $J^*_{\text{app}}(k)$ at each event step k, the initial state we use for the MPC problem for step k is the state $x(k)$ obtained from the simulation using the numerical integration approach. This figure also shows the upper bound for $J^*_{\text{app}}(k) - J^*(k)$, which is obtained by using (12). As shown in Figure 3, after a few steps, for $p = 20, 30$, and 40 the upper bound for $J^*_{\text{app}}(k) - J^*(k)$ is below 1. However, for $p = 100$ both the upper bound for $J^*_{\text{app}}(k) - J^*(k)$ and the difference itself become larger than 2. Experiments (not reported here) using the 5σ- and 6σ-rule instead of the 3σ-rule for defining L show that the approximation error due to inequality (*i*) of Proposition 2 is negligible. Hence, we now focus on the error due to (*ii*) and the upper bound for the error of Jensen's inequality due to (*iii*), as given in (13). As shown in Figure 3, this upper bound also become closer to the exact solution after few steps and by increasing p, the difference between this upper bound and the exact solution becomes larger.

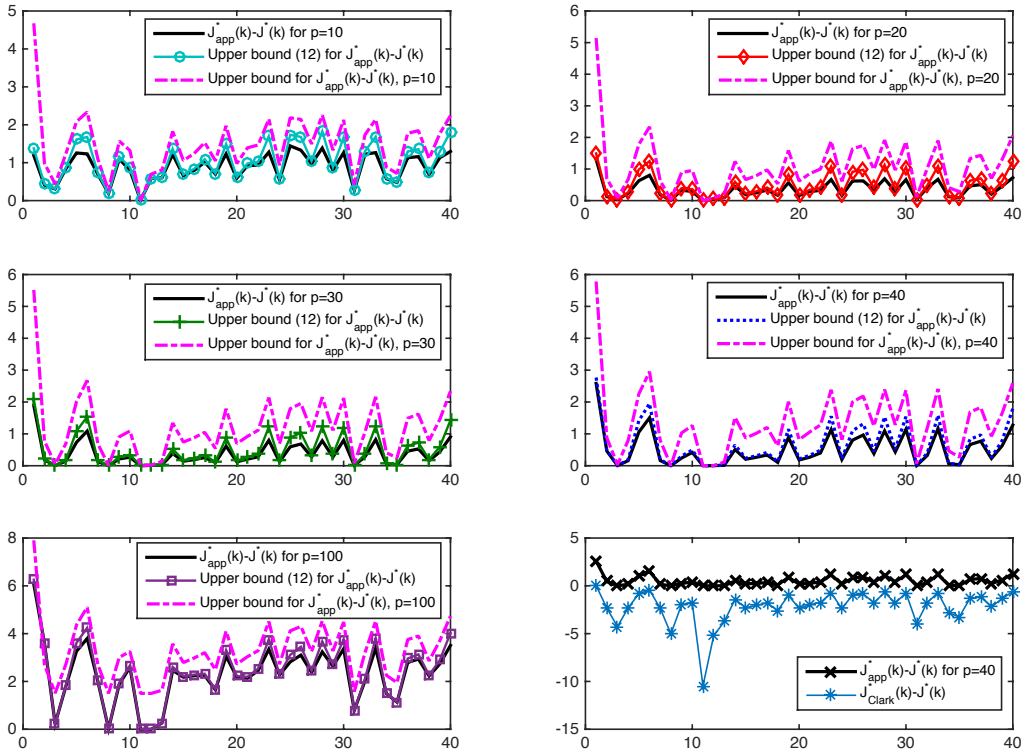

Figure 3: The upper bound (12) for and the exact value of $J^*_{\text{app}}(k) - J^*(k)$ for different values of p.

These observations support our previous conclusion that a larger p is not always a better option and hence, p has to be chosen such that $J^*_{\text{app}}(k) - J^*(k)$ and accordingly, the upper bound for this difference have the possible smallest value. The last plot in Figure 3 shows the difference between the exact solution and the approximation method with $p = 40$ and between the exact solution and the approximation method proposed in [12], denoted by $J^*_{\text{Clark}}(k)$. As explained in Remark 7 and as can be seen here, the method of [12] does not provide an upper bound for the expected value of maximum of random variables and hence, for the objective function. Actually, in this example, this method provides a lower bound for the exact solution, which means that by minimizing the objective function using this approximation method, we get further away from the exact solution. Also, due to the iterative nature of the method of [12], the CPU time of the closed-loop MPC optimization using this approach is 78.57 s, which is approximately 2.5 times larger than the one (i.e., 30.63 s) using the proposed approximation method with $p = 40$.

Consequently, based on the computation time, the mean value of J_{tot}, and the relative and the approximation errors, we can conclude that the approximation method is a reliable and time-efficient method to solve the stochastic max-plus-linear MPC optimization problem in this ex-

ample. Indeed, for $p = 20,\ 30$, and 40 the result of the closed-loop optimization using the approximation method is close to the one using numerical integration and, in this specific example, the approximation-based approach is about 30 times faster than the approach using numerical integration.

5. Conclusions

We have introduced an approximation method to reduce the computational complexity imposed by the computation of the expected value of max-affine expressions, which appears in optimization problems for different types of stochastic discrete-event and hybrid systems, such as stochastic max-plus-linear systems or stochastic monotonic PWA systems. This expected value usually appears in the objective function and/or in the constraints of the optimization problem and creates a significant computational load. To tackle this issue, we have proposed an approximation method that is based on the moments of stochastic variables. We assume that the random variables in the system are independent and as an example they are assumed to be normally distributed. Since a closed-form expression of the moments of a normally distributed random variable exists, we can approximate the expected value of max-affine expressions without using numerical integration. It is important to note that this method is not limited to normally distributed random variables, but that it is in general applicable to any distribution that is preserved under summation and that has finite moments for which a closed-form expression exists, such as the Poisson distribution and the gamma distribution. It is also applicable to random variables that their sum has a known distribution with finite moments for which a closed-form expression exists, such as i.i.d random variables with uniform distribution $U(0, 1)$ the sum of which has the Irwin-Hall distribution with finite moments and closed-form expressions. We have also proved that the approximation function, and consequently, the cost function are convex in the control (optimization) variable. Therefore, if the constraints are non-decreasing affine functions of the output, the resulting optimization problem is convex, which thus can be solved efficiently.

As a particular application of this approximation method, we applied it to the MPC optimization problem of max-plus-linear systems. MPC is a very popular control approach for practical applications; therefore, the proposed approximation method might be useful to enable and promote the use of max-plus MPC methods in the control of production processes. Moreover, we have shown in this example that by choosing the appropriate order of moments, we can decrease the approximation error (for the given example). In this paper, this choice was made by means of numerical experiments. However, it is a topic for future research to find more efficient ways to determine the appropriate order of moments. In addition, the system we have studied in this example was quite simple; hence, in the future research, we will obtain more efficient implementations and/or additional approximation methods to improve the scalability of the current approach and to apply it to more complex and large scale systems. Moreover, it is interesting to find out how large the approximate error will be when evaluating large sets of affine expressions, occurring for multivariate systems with many components or for MPC with a long time horizon. Another topic for the future research is to extend this approximation such that it is applicable to a wider range of distributions and also to find a tighter upper bound for the error of this approximation method. Yet another topic for future work would be the extensive assessment and comparison of this method with other approximation methods such as the variability expansion method of [53] and approximation methods of [36, 37].

6. Appendix: MPC for stochastic max-plus linear systems

6.1. Problem description

In [17, 52] the MPC framework has been extended to max-plus-linear models

$$x(k) = A(k) \otimes x(k-1) \oplus B(k) \otimes u(k) \tag{19}$$

$$y(k) = C(k) \otimes x(k) \tag{20}$$

where $A(k)$, $B(k)$, and $C(k)$ are systems matrices, $x(k)$ is the state at event step k, and $u(k)$ and $y(k)$ are the input and the output [3]. In fact, $x(k)$, $u(k)$, and $y(k)$ contain the time instants at which the internal, input, and output events occur for the k-th time, respectively. We define a cost criterion

$$J(x(k-1), \tilde{u}(k)) = \underbrace{\sum_{j=0}^{N_\mathrm{p}-1} \sum_{i=1}^{n_y} \mathbb{E}[\kappa_i(k+j)]}_{J_\mathrm{out}(x(k-1),\tilde{u}(k))} + \lambda \underbrace{\Big(- \sum_{j=0}^{N_\mathrm{p}-1} \sum_{l=1}^{n_u} u_l(k+j) \Big)}_{J_\mathrm{in}(\tilde{u}(k))} \tag{21}$$

where N_p is the prediction horizon, $\lambda > 0$ is a weighting factor, $\mathbb{E}[\cdot]$ denotes the expected value operator, $\kappa_i(k) = \max(y_i(k) - r_i(k), 0)$ penalizes the late (but not the early) deliveries for the i-th output at event step k where $r(k)$ is the vector of due date signals, and $\tilde{u}(k) = [\, u^T(k) \;\; u^T(k+1) \;\; \cdots \;\; u^T(k+N_\mathrm{p}-1) \,]^T$ ($\tilde{y}(k)$ is defined similarly). The stochastic max-plus-linear MPC problem at event step k is now defined as [17, 52]

$$\begin{aligned} \min_{\tilde{u}(k),\tilde{y}(k),\tilde{x}(k)} \; & J_\mathrm{out}(x(k), \tilde{u}(k)) + \lambda J_\mathrm{in}(\tilde{u}(k)) \\ \text{s.t.} \;\; & (19) \;\text{and}\; (20) \\ & u(k+j) \geq u(k+j-1), \quad j = 0, \ldots, N_\mathrm{p} - 1 \\ & A_\mathrm{con}(k)\tilde{u}(k) + B_\mathrm{con}(k)\mathbb{E}[\tilde{y}(k)] \leq c_\mathrm{con}(k) \end{aligned} \tag{22}$$

where $A_\mathrm{con}(k)$ and $B_\mathrm{con}(k)$ are the linear constraint matrices. For more details, we refer the interested reader to [17, 52]. If all entries of $B_\mathrm{con}(k)$ are nonnegative, the stochastic max-plus-linear MPC problem (22) is a convex problem [52, Property 4].

Let $\tilde{\kappa}(k)$, $\tilde{r}(k)$, and $\tilde{e}(k)$ be defined in a similar way as $\tilde{u}(k)$. To solve the max-plus-linear MPC optimization problem (22), we need to compute the expected value of the signals $\tilde{\kappa}(k)$ and $\tilde{y}(k)$. As shown in [52, Lemma 2], the entries of both $\tilde{\kappa}(k)$ and $\tilde{y}(k)$ belong to $\mathcal{S}_\mathrm{ma}(z(k))$ (defined in Section 2) with $z(k) = [-\tilde{r}^T(k) \;\; x^T(k-1) \;\; \tilde{u}^T(k) \;\; \tilde{e}^T(k)]^T$. Now, let us rewrite both $\tilde{\kappa}(k)$ and $\tilde{y}(k)$ as a general max-affine function of $\tilde{u}(k)$ and $\tilde{e}(k)$ [52], and denote it by $v(k)$ as follows:

$$v(k) = \max_{j=1,\ldots,n_v} (\xi_j + \beta_j^T \tilde{u}(k) + \delta_j^T w(k) + \gamma_j^T \tilde{e}(k)) \tag{23}$$

where n_v is the number of terms that appear in the maximization, $\xi_j \in \mathbb{R}_\varepsilon$, $\beta_j \in (\mathbb{R}^+)^{n_u}$, $\delta_j \in (\mathbb{R}^+)^{n_w}$, $\gamma_j \in (\mathbb{R}^+)^{n_{\tilde{e}}}$, with $w(k) = [-\tilde{r}^T(k) \;\; x^T(k-1)]^T$ the vector of non-stochastic variables, and $\tilde{e}(k) \in \mathbb{R}^{n_{\tilde{e}}}$ a stochastic variable with probability density function $f(\cdot)$.

For a shorter notation let $\alpha_j(k) = \xi_j + \delta_j^T w(k) + \beta_j^T \tilde{u}(k)$. Hence, we get

$$v(k) = \max_{j=1,\ldots,n_v} (\alpha_j(k) + \gamma_j^T \tilde{e}(k)) \tag{24}$$

which is a convex max-affine expression in $\tilde{e}(k)$. Accordingly, we can solve the optimization problem (22) by means of the approximation method proposed in Section 2 of this paper.

6.2. Alternative approaches

To solve the optimization problem (22), the expected value of $v(k)$ has to be calculated, which is defined as follows:

$$\begin{aligned}\mathbb{E}[v(k)] &= \int_{-\infty}^{+\infty} \cdots \int_{-\infty}^{+\infty} v(k) f(\tilde{e}) d\tilde{e} \\ &= \sum_{j=1}^{n_v} \int_{\tilde{e} \in \Phi_j(\tilde{u}(k))} \cdots \int (\alpha_j(k) + \gamma_j^T \tilde{e}) f(\tilde{e}) d\tilde{e} \end{aligned} \tag{25}$$

where $d\tilde{e} = d\tilde{e}_1 d\tilde{e}_2 \dots d\tilde{e}_{n_{\tilde{e}}}$ and the sets $\Phi_j(\tilde{u}(k))$ are defined such that for $j = 1, \dots, n_v$ we have $v(k) = \alpha_j(k) + \gamma_j^T \tilde{e}(k)$ for all $\tilde{e} \in \Phi_j(\tilde{u}(k))$ and $\bigcup_{j=1}^{n_v} \Phi_j(\tilde{u}(k)) = \mathbb{R}^{n_{\tilde{e}}}$. Note that the sets Φ_j constitute a partition of $\mathbb{R}^{n_{\tilde{e}}}$. It is shown in [52, Proposition 3] that the function $\mathbb{E}[v(k)]$ is convex in $\tilde{u}(k)$.

One way of computing $\mathbb{E}[v(k)]$ in (25) is to use numerical integration. The common methods for numerical integration are (non)adaptive integration, (non)iterative integration, exponential quadrature, Monte Carlo integration, the Nyström method, the Quasi-Monte Carlo method, and the multi-step method [15]. However, numerical integration is in general both cumbersome and time-consuming, and it becomes even more complicated as the probability density function f becomes more and more complex.

In [52], an alternative method for computing $\mathbb{E}[v(k)]$ is proposed based on analytic integration. To that end, a piecewise polynomial probability density function defined on polyhedral sets is considered. Such a function can be obtained in two ways: either the stochastic vector already has a piecewise polynomial probability density function (such as the uniform distribution) or we approximate the real probability density function with a piecewise polynomial probability density function[11] (such as the normal distribution where its probability density function can be approximated by PWA functions). Accordingly, $\mathbb{E}[v(k)]$ can be written as

$$\mathbb{E}[v(k)] = \sum_{\ell=1}^{n_p} \sum_{j=1}^{n_v} \int_{\tilde{e} \in \Psi_{j\ell}(\tilde{u}(k))} \cdots \int (\alpha_j(k) + \beta_j^T \tilde{u}(k) + \gamma_j^T \tilde{e}) f_\ell(\tilde{e}) d\tilde{e}. \tag{26}$$

where $f_\ell(\tilde{e})$ is the piecewise polynomial probability density function of $\tilde{e}$. The expression (26) is a sum of integrals of polynomial functions in $\tilde{e}$ and hence, can be solved analytically for each polyhedron $\Psi_{ij\ell}$ [9, 31].

Even if the integral in (26) can be computed analytically, the computational load is still quite heavy. This is due to the fact that this method contains two time-consuming steps: In the first step all polyhedra $\Psi_{j\ell}$ have to be specified, where the number of polyhedra Φ_j is equal to n_v and the number of polyhedra P_ℓ is n_p. Hence, in the worst case the number of polyhedra $\Psi_{j\ell}$ that has to be considered is $O(n_v n_p)$, which becomes more and more time-consuming as n_p and n_v become larger. In the second step, the integral over each of these regions has to be calculated, for which in the simplest case of having a uniform probability density function, we need to compute all the vertices of each polyhedron $\Psi_{j\ell}$. As explained in [34], we have the following upper bound

[11] The approximate probability density function must be nonnegative and its integral over the domain of the real probability density function must be equal to 1. This can be assured by including these conditions as constraints in the parameter fitting.

for the number of the vertices of a polytope defined by n_v (non-redundant) inequality constraints in an $n_{\tilde{e}}$-dimensional space:

$$\mathcal{V}(n_v, n_{\tilde{e}}) = \binom{n_v - \lfloor \frac{n_{\tilde{e}}+1}{2} \rfloor}{n_v - n_{\tilde{e}}} + \binom{n_v - \lfloor \frac{n_{\tilde{e}}+2}{2} \rfloor}{n_v - n_{\tilde{e}}} \tag{27}$$

This means that in our case, where typically n_v is much larger than $n_{\tilde{e}}$, i.e., $n_v \gg n_{\tilde{e}} \gg 1$, the number of vertices for the worst case can be $O(n_v^{\lfloor \frac{n_{\tilde{e}}}{2} \rfloor})$, which is again time-consuming as n_v and $n_{\tilde{e}}$ increase. Accordingly for the case of a uniformly distributed noise, the complexity of the whole procedure in the worst case is of order $O(n_p n_v^{\lfloor \frac{n_{\tilde{e}}}{2} \rfloor + 1})$. In the case of other piecewise polynomial probability density functions, the order of complexity of the second step becomes even bigger since then, the integral computation is more complex than the one in the case of a uniform distribution. Therefore, the computational complexity of this method increases exponentially as $n_{\tilde{e}}$ increases and polynomially as n_v increases. It increases even more in the case of non-piecewise polynomial probability density functions, such as a normal probability density function, since these functions cannot be directly applied for this method and have to be approximated by piecewise polynomial probability density functions.

In [53] an effort is made to reduce the computational complexity of the above-mentioned method by approximating $\mathbb{E}[v(k)]$ in (25) using the method of variability expansion. Since variability expansion is an analytic method and does not resort to simulation, it is, in principle, possible to compute higher-order moments of performance characteristics of stochastic systems. As presented in [53, Section 4], it is assumed that the entries of $\tilde{e}(k)$ are independent and identically distributed (i.i.d) and an artificial parameter $\theta \in [0, 1]$ is introduced. The i-th entry of $\tilde{e}(k)$ is then replaced by its mean with probability $1 - \theta$ and the result is denoted by $\tilde{e}_\theta(k)$. The parameter θ allows controlling the level of randomness in the system, and letting θ go from 0 to 1 increases the level of stochasticity in the system. The main idea of variability expansion is as follows: considering $\mathbb{E}[v_\theta(k)]$ as a function of θ, it can be developed into a Taylor series in θ that converges to the true function on some subset $\mathcal{X} \subseteq \mathbb{R}$. In particular, if the value of $d^m/d\theta^m \mathbb{E}[v_\theta(k)]$ for $\theta = 0$ is denoted by $d^m/d\theta^m \mathbb{E}[v_0(k)]$, then $\mathbb{E}[v_1(k)]$, the "true" expected value of $v(k)$ is given by

$$\mathbb{E}[v(k)] = \mathbb{E}[v_1(k)] = \sum_{m=0}^{M} \frac{1}{m!} \frac{d^m}{d\theta^m} \mathbb{E}[v_0(k)] + R_M(k) \tag{28}$$

where for $M < n_{\tilde{e}}$

$$R_M(k) \leq \frac{1}{(M+1)!} \sup_{\theta \in [0,1]} \left| \frac{d^{M+1}}{d\theta^{M+1}} \mathbb{E}[v_0(k)] \right|$$

and $R_M = 0$ otherwise [24]. It is been also shown in [53] that a closed-form expression for the m-th order derivative $d^m/d\theta^m \mathbb{E}[v_0(k)]$ can be obtained.

The computational complexity of approximating $\mathbb{E}[v(k)]$ using the method of variability expansion has been discussed in [53, Section 5]. Based on this discussion, the overall complexity will at least be of order

$$O\Big(\binom{n_{\tilde{e}}}{M} n_v \mathcal{V}(2M + n_v - 1, M)\Big) = O\Big(\frac{n_{\tilde{e}}^M \; n_v^{\frac{M+1}{2}}}{M! \; (\frac{M-1}{2})!}\Big)$$

where $\mathcal{V}(\cdot, \cdot)$ is given in (27). The derivation of the above error order can be found in Appendix A. Clearly, the computational complexity increases polynomially if n_e and n_v increase and exponentially if M increases.

References

[1] M.A. Abramowitz, I. Stegun, Handbook of Mathematical Functions, National Bureau of Standards, US Government Printing Office, Washington DC, 1964.

[2] M. Akian, Representation of stationary solutions of Hamilton-Jacobi-Bellman equations: a max-plus point of view, in: Proceedings of the International Workshop "Idempotent and Tropical Mathematics and Problems of Mathematical Physics", Moscow, Russia.

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

[4] A. Bemporad, M. Morari, Control of systems integrating logic, dynamics, and constraints, Automatica 35 (1999) 407–427.

[5] W.H. Beyer, CRC Standard Mathematical Tables, CRC Press, Boca Raton, Florida, 28th edition, 1987.

[6] J.L. Boimond, J.L. Ferrier, Internal model control and max-algebra: Controller design, IEEE Transactions on Automatic Control 41 (1996) 457–461.

[7] J.L. Boimond, L. Hardouin, 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.

[8] S. Boyd, L. Vandenberghe, Convex Optimization, Cambridge University Press, London, UK, 2004.

[9] B. Büeler, A. Enge, K. Fukuda, Exact volume computation for convex polytopes: A practical study, in: G. Kalai, G. Ziegler (Eds.), Polytopes – Combinatorics and Computation, Birkäuser Verlag, Basel, Switzerland, 2000, pp. 131–154.

[10] E.F. Camacho, C. Bordons, Model Predictive Control in the Process Industry, Springer-Verlag, Berlin, Germany, 1995.

[11] L.O. Chua, A.C. Deng, Canonical piecewise-linear representation, IEEE Transactions on Circuits and Systems 35 (1988) 101–111.

[12] C.E. Clark, The greatest of a finite set of random variables, Operations Research 9 (1961) 145–162.

[13] B. Cottenceau, L. Hardouin, J.L. Boimond, J.L. Ferrier, Model reference control for timed event graphs in dioids, Automatica 37 (2001) 1451–1458.

[14] R.A. Cuninghame-Green, Minimax Algebra, Springer-Verlag, Berlin, Germany, 1979.

[15] P.J. Davis, P. Rabinowitz, Methods of Numerical Integration, Academic Press, New York, 2nd edition, 1984.

[16] B. De Schutter, B. De Moor, Minimal realization in the max algebra is an extended linear complementarity problem, Systems & Control Letters 25 (1995) 103–111.

[17] B. De Schutter, T. van den Boom, Model predictive control for max-plus-linear discrete event systems, Automatica 37 (2001) 1049–1056.

[18] B. De Schutter, T. van den Boom, V. Verdult, 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, Las Vegas, Nevada, pp. 4024–4029.

[19] F.M. Dekking, C. Kraaikamp, H.P. Lopuhaä, L.E. Meester, A Modern Introduction to Probability and Statistics, Springer, London, UK, 2005.

[20] F. Gallot, J.L. Boimond, L. Hardouin, Identification of simple elements in max-algebra: Application to SISO discrete event systems modelisation, in: Proceedings of the European Control Conference (ECC'97), Brussels, Belgium, paper 488.

[21] C.E. García, D.M. Prett, M. Morari, Model predictive control: Theory and practice — A survey, Automatica 25 (1989) 335–348.

[22] G.H. Golub, C.F. Van Loan, Matrix Computations, The John Hopkins University Press, Baltimore, Maryland, 3rd edition, 1996.

[23] W.P.M.H. Heemels, B. De Schutter, A. Bemporad, Equivalence of hybrid dynamical models, Automatica 37 (2001) 1085–1091.

[24] B. Heidergott, Max-Plus Linear Stochastic Systems and Perturbation Analysis, Springer, New York, 2006.

[25] B. Heidergott, G.J. Olsder, J. van der Woude, Max Plus at Work, Princeton University Press, Princeton, New Jersey, 2006.

[26] M. Johansson, Piecewise Linear Control Systems, volume 284 of *Lecture Notes in Control and Information Sciences*, Springer, Berlin, Germany, 2003.

[27] M. Johansson, A. Rantzer, Computation of piecewise quadratic Lyapunov functions for hybrid systems, IEEE Transactions on Automatic Control 43 (1998) 555–559.

[28] N.L. Johnson, S. Kotz, N. Balakrishnan, Continuous Univariate Distributions, Vol. 2, Wiley, New York, 2nd edition, 1995.

[29] T.A.M. Kevenaar, D.M.W. Leenaerts, A comparison of piecewise-linear model descriptions, IEEE Transactions on Circuits and Systems – I: Fundamental Theory and Applications 39 (1992) 996–1004.

[30] N.K. Krivulin, The max-plus algebra approach in modelling of queueing networks, in: Proceedings of the 1996 Summer Computer Simulation Conference, San Diego, California, pp. 485–490.
[31] J.B. Lasserre, Integration on a convex polytope, Proceedings of the American Mathematical Society 126 (1998) 2433–2441.
[32] J.M. Maciejowski, Predictive Control with Constraints, Prentice Hall, Harlow, England, 2002.
[33] J. Mairesse, Stochastic linear systems in the (max,+) algebra, in: 11th International Conference on Analysis and Optimization of Systems Discrete Event Systems, volume 199 of *Lecture Notes in Control and Information Sciences*, Springer-Verlag, London, UK, 1994, pp. 257–265.
[34] T.H. Mattheiss, D.S. Rubin, A survey and comparison of methods for finding all vertices of convex polyhedral sets, Mathematics of Operations Research 5 (1980) 167–185.
[35] W.M. McEneaney, Max-plus eigenvector methods for nonlinear H-infinity problems: Error analysis, SIAM Journal of Control and Optimization 43 (2004) 379–412.
[36] W.M. McEneaney, Max-Plus Methods for Nonlinear Control and Estimation, Systems and Control Series, Birkhauser, 2006.
[37] W.M. McEneaney, A. Deshpande, S. Gaubert, Curse-of-complexity attenuation in the curse-of-dimentionality-free method for hjb pdes, in: Proceedings of the Amercian Control Conference.
[38] E. Menguy, J.L. Boimond, L. Hardouin, J.L. Ferrier, A first step towards adaptive control for linear systems in max algebra, Discrete Event Dynamic Systems 10 (2000) 347–368.
[39] D.S. Mitrinović, J.E. Pečrić, A.M. Fink, Classical and New Inequalities in Analysis, Kluwer Academic Publishers, Dordrecht, The Netherlands, 1993.
[40] I. Necoara, T.J.J. van den Boom, B. De Schutter, J. Hellendoorn, Stable model predictive control for constrained max-plus-linear systems, Discrete Event Dynamic Systems: Theory and Applications 17 (2007) 329–354.
[41] Y. Nesterov, A. Nemirovskii, Interior-Point Polynomial Algorithms in Convex Programming, SIAM, Philadelphia, Pennsylvania, 1994.
[42] G.J. Olsder, J.A.C. Resing, R.E. de Vries, M.S. Keane, G. Hooghiemstra, Discrete event systems with stochastic processing times, IEEE Transactions on Automatic Control 35 (1990) 299–302.
[43] A. Papoulis, Probability, Random Variables, and Stochastic Processes, McGraw-Hill, Singapore, 1991.
[44] J.A.C. Resing, R.E. de Vries, G. Hooghiemstra, M.S. Keane, G.J. Olsder, Asymptotic behavior of random discrete event systems, Stochastic Processes and Their Applications 36 (1990) 195–216.
[45] G. Schullerus, V. Krebs, Diagnosis of batch processes based on parameter estimation of discrete event models, in: Proceedings of the European Control Conference 2001 (ECC'01), Porto, Portugal, pp. 1612–1617.
[46] G. Schullerus, 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, Seoul, Korea.
[47] G. Schullerus, V. Krebs, B. De Schutter, T. van den Boom, Input signal design for identification of max-plus-linear systems, Automatica 42 (2006) 937–943.
[48] S. Simić, On a new converse of Jensen's inequality, Publications de L'Institut Mathématique 85 (2009) 107–110.
[49] K.K. Somasundaram, J.S. Baras, Solving multi-metric network problems an interplay between idempotent semiring rules, Linear Algebra and its Applications 435 (2011) 1494–1512.
[50] E.D. Sontag, Nonlinear regulation: The piecewise linear approach, IEEE Transactions on Automatic Control 26 (1981) 346–358.
[51] T. van den Boom, B. De Schutter, Properties of MPC for max-plus-linear systems, European Journal of Control 8 (2002) 453–462.
[52] T. van den Boom, B. De Schutter, Model predictive control for perturbed max-plus-linear systems: A stochastic approach, International Journal of Control 77 (2004) 302–309.
[53] T.J.J. Van den Boom, B. Heidergott, B. De Schutter, Complexity reduction in MPC for stochastic max-plus-linear discrete event systems by variability expansion, Automatica 43 (2007) 1058–1063.
[54] L. Vandenberghe, B. De Moor, J. Vandewalle, The generalized linear complementarity problem applied to the complete analysis of resistive piecewise-linear circuits, IEEE Transactions on Circuits and Systems 36 (1989) 1382–1391.
[55] R. Willink, Normal moments and Hermite polynomials, Statistics & Probability Letters 73 (2005) 271–275.
[56] S.J. Wright, Primal-Dual Interior Point Methods, SIAM, Philadelphia, Pennsylvania, 1997.