Delft University of Technology

Delft Center for Systems and Control

Technical report 10-039

An approximation approach for model predictive control of stochastic max-plus linear systems*

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 approach for model predictive control of stochastic max-plus linear systems," *Proceedings of the 10th International Workshop on Discrete Event Systems (WODES 2010)*, Berlin, Germany, pp. 376–381, Aug.–Sept. 2010. doi:10.3182/20100830-3-DE-4013.00062

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/10_039.html

An Approximation Approach for Model Predictive Control of Stochastic Max-Plus Linear Systems ⋆

Samira S. Farahani * **Ton van den Boom** * **Hans van der Weide** ** **Bart De Schutter** *

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

Abstract: Model Predictive Control (MPC) is a model-based control method based on a receding horizon approach and online optimization. In previous work we have extended MPC to a class of discrete-event systems, namely the max-plus linear systems, i.e., models that are "linear" in the max-plus algebra. Lately, the application of MPC for stochastic max-plus-linear systems has attracted a lot of attention. At each event step, an optimization problem then has to be solved that is, in general, a highly complex and computationally hard problem. Therefore, the focus of this paper is on decreasing the computational complexity of the optimization problem. To this end, we use an approximation approach that is based on the p-th raw moments of a random variable. This method results in a much lower computational complexity and computation time while still guaranteeing a good performance.

Keywords: Stochastic discrete event systems, stochastic max-plus linear systems, model predictive control, approximation, raw moments

1. INTRODUCTION

Model predictive control (MPC) (Rawlings and Mayne, 2009) is an advanced control approach that relies on a dynamic model of the process and is capable of handling constraints on inputs or outputs in a systematic way. Although conventional MPC uses linear or nonlinear discrete-time model, MPC has also been extended to discrete-event systems (De Schutter and van den Boom, 2001; Necoara et al., 2004).

The class of discrete-event systems essentially consists of man-made systems that contain a finite number of resources (such as machines, communications channels, or processors) that are shared by several users (such as product types, information packets, or jobs) all of which contribute to the achievement of some common goal (the assembly of products, the end-to-end transmission of a set of information packets, or a parallel computation) (Baccelli et al., 1992). In this paper we consider a special class of discrete-event systems, namely the max-plus linear (MPL) systems. Loosely speaking, this class contains discrete event systems with synchronization but no choice. Models of such systems are based on two main operations, maximization and addition. This leads to a description that is linear in max-plus algebra (Baccelli et al., 1992; Cuninghame-Green, 1979; Heidergott et al., 2006), and that applies to both deterministic and stochastic discrete-event systems. Some results for the analysis of stochastic MPL discrete-event systems can be found in (Olsder et al., 1990; Resing et al., 1990).

One of the relevant topics that has attracted much attention recently, is the application of MPC for perturbed max-plus-linear systems in which modeling errors, noise, and/or disturbances are present. Van den Boom and De Schutter (2004) have developed an MPC controller for such perturbed MPL systems. They also showed that the resulting MPC optimization problem is convex under quite general conditions. However, by increasing the prediction horizon and the system order, the computational complexity increases significantly. In (Van den Boom et al., 2007) an effort is made to reduce the complexity by introducing an approximation method, namely the variability expansion. However, despite the resulting complexity reduction, the level of the complexity of the main problem remains too high.

This paper focuses on another approach to reduce the complexity of MPC for perturbed MPL systems by approximating the calculation of stochastic integrals using raw moments of random variables. This method simplifies the computations considerably and by choosing the appropriate order p of the raw moments, the approximation error can be made sufficiently small. Moreover, due to the special structure of the p-th raw moments, this method also results in a convex optimization problem that can be solved efficiently. Since we can compute these moments analytically, this approach results in a much faster and more efficient way to solve the stochastic MPL-MPC problem without increasing the computational complexity.

The paper is organized as follows. Section 2 provides some background information on max-plus algebra and stochastic MPL systems. In Section 3 we give a concise account of the MPC algorithm for stochastic MPL systems presented in (Van den Boom and De Schutter, 2004). Section 4 introduces the new approach based on p-th raw moments and describes how it reduces the complexity of the MPC optimization problem. In Section 5, we study the convexity of the MPC MPL problem after applying the approximation method. Section 6 presents a worked example in which the computation time of the ap-

⋆ Research supported by the EU project "Highly-complex and networked control systems (HYCON2)" and by the STW project "Model-predictive railway traffic management – A framework for closed-loop control of large-scale railway systems".

proximation method is compared with the one from the analytic computation.

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

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

2.1 Max-Plus Algebra

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

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

for $x, y \in \mathbb{R}_\varepsilon$. Note that the zero element of the max-plus addition is ε, i.e., $x \oplus \varepsilon = x$, and the identity element of the max-plus multiplication is $e = 0$, i.e., $x \otimes e = x$. The corresponding max-plus matrix operations are defined as

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

for $A, B \in \mathbb{R}_\varepsilon^{m \times n}$ and $C \in \mathbb{R}_\varepsilon^{n \times p}$.

2.2 Max-Plus-Nonnegative-Scaling Functions

Let $\mathscr{S}_{\text{mpns}}$ denote the set of max-plus-nonnegative-scaling functions, i.e., functions f of the form

$$f(z) = \max_{i=1,\ldots,m} (\tau_{i,1} z_1 + \cdots + \tau_{i,n} z_n + \xi_i)$$

with variable $z \in \mathbb{R}_\varepsilon^n$ and constant coefficients $\tau_{i,j} \in \mathbb{R}^+$ and $\xi_i \in \mathbb{R}$, where $\mathbb{R}^+$ is the set of the nonnegative real numbers. In the sequel, we stress that f is a function of z by writing $f \in \mathscr{S}_{\text{mpns}}(z)$. As shown by Van den Boom and De Schutter (2004), the set $\mathscr{S}_{\text{mpns}}$ is closed under the operations $\oplus, \otimes$, and the scalar multiplication by a nonnegative scalar.

2.3 Stochastic MPL Systems

Discrete event systems with synchronization but no choice can be modeled as follows (Baccelli et al., 1992; Cuninghame-Green, 1979):

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

where $x(k)$ is the state of the system at event step k, and $u(k)$ and $y(k)$ are the input and output of the system. In fact, the vectors $u(k)$ and $y(k)$ contain the time instants at which the internal input and output events occur for the k-th time, respectively. Since in a stochastic system noise leads to perturbations of the system matrices, $A(k)$, $B(k)$, and $C(k)$ are in general uncertain system matrices. Following Van den Boom and De Schutter (2004) and Van den Boom et al. (2007), these uncertainties are presented by the vector $e(k)$, which is a stochastic variable with a certain probability distribution. Hence, the entries of the system matrices belong to $\mathscr{S}_{\text{mpns}}$ (Van den Boom and De Schutter, 2004), i.e., $A(k) \in \mathscr{S}_{\text{mpns}}^{n \times n}(e(k))$, $B(k) \in \mathscr{S}_{\text{mpns}}^{n \times n_u}(e(k))$, $C(k) \in \mathscr{S}_{\text{mpns}}^{n_y \times n}(e(k))$.

3. MPC FOR STOCHASTIC MPL SYSTEMS

In (De Schutter and van den Boom, 2001; Van den Boom and De Schutter, 2004) the MPC framework has been extended to MPL models (1)–(2) as follows. Following the conventional MPC methodology (Rawlings and Mayne, 2009), we define a cost criterion J that reflects the input and output cost functions in the event period $[k, k+N_\text{p}-1]$:

$$J(k) = J_{\text{out}}(k) + \lambda J_{\text{in}}(k)$$

where N_p is the prediction horizon and λ is a weighting factor. Since we consider a stochastic system, the cost criterion is actually defined as

$$J(k) = \sum_{j=0}^{N_\text{p}-1} \sum_{i=1}^{n_y} \mathbb{E}[\kappa_i(k+j)] - \lambda \sum_{j=0}^{N_\text{p}-1} \sum_{l=1}^{n_u} u_l(k+j) \tag{3}$$

where $\mathbb{E}[\cdot]$ denotes the expected value operator, $\kappa_i(k) = \max(y_i(k) - r_i(k), 0)$ is the tardiness error for the i-th output at event step k, and $r(k)$ is the vector of reference (due date) signals. The aim is to compute an optimal input sequence $u(k), \ldots, u(k+N_\text{p}-1)$ that minimizes $J(k)$ subject to linear constraints on the inputs and outputs as discussed in Van den Boom and De Schutter (2004). Since the $u(k)$'s correspond to consecutive event occurrence times, we have the additional condition $\Delta u(k+j) = u(k+j) - u(k+j-1) \geq 0$ for $j = 0, \ldots, N_\text{p}-1$. Furthermore, in order to reduce the number of decision variables and the corresponding computational complexity we introduce a control horizon N_c ($\leq N_\text{p}$) and we impose the additional condition that the input rate should be constant from the point $k+N_\text{c}-1$ on: $\Delta u(k+j) = \Delta u(k+N_\text{c}-1)$ for $j = N_\text{c}, \ldots, N_\text{p}-1$, or equivalently $\Delta^2 u(k+j) = \Delta u(k+j) - \Delta u(k+j-1) = 0$ for $j = N_\text{c}, \ldots, N_\text{p}-1$. MPC uses a receding horizon principle which means that after computation of the optimal control sequence $u(k), \ldots, u(k+N_\text{c}-1)$, only the first control sample $u(k)$ will be implemented, subsequently the horizon is shifted one event step, and the optimization is restarted with new information of the measurements. Consider the following vectors:

$$\tilde{u}(k) = [\ u^T(k) \ \cdots \ u^T(k+N_\text{p}-1)\ ]^T,$$
$$\tilde{r}(k) = [\ r^T(k) \ \cdots \ r^T(k+N_\text{p}-1)\ ]^T,$$
$$\tilde{y}(k) = [\ y^T(k) \ \cdots \ y^T(k+N_\text{p}-1)\ ]^T,$$
$$\tilde{e}(k) = [\ e^T(k) \ \cdots \ e^T(k+N_\text{p}-1)\ ]^T.$$

By using successive substitution on (1)-(2), we get $\tilde{y}(k) = \tilde{C}(\tilde{e}(k)) \otimes x(k-1) \oplus \tilde{D}(\tilde{e}(k)) \otimes \tilde{u}(k)$ for appropriately defined matrices $\tilde{C}(\tilde{e}(k))$ and $\tilde{D}(\tilde{e}(k))$ (Van den Boom and De Schutter, 2004). The stochastic MPL-MPC problem for event step k is then defined as follows (Van den Boom and De Schutter, 2004):

$$\min_{\tilde{u}(k)} J_{\text{out}}(k) + \lambda J_{\text{in}}(k)$$
$$\text{s.t.} \Delta u(k+j) \geq 0 \quad \text{for } j = 0, \ldots, N_\text{p}-1$$
$$\Delta^2 u(k+j) = 0 \quad \text{for } j = N_\text{c}, \ldots, N_\text{p}-1$$
$$A_{\text{con}}(k)\tilde{u}(k) + B_{\text{con}}(k)\mathbb{E}[\tilde{y}(k)] \leq c_{\text{con}}(k)$$

where the last inequality is in fact a soft constraint due to the expected value of $y(k)$ (so $c_{\text{con}}(k)$ will in general include a safety margin if necessary).

To solve the above optimization problem, we need to compute the expected value of the signals $\tilde{\kappa}(k)$ and $\tilde{y}(k)$. By Lemma 2 of (Van den Boom and De Schutter, 2004), both $\tilde{\kappa}(k)$ and $\tilde{y}(k)$ belong to $\mathscr{S}_{\text{mpns}}(z(k))$ with $z(k) = [-\tilde{r}^T(k)\ x^T(k-1)\ \tilde{u}^T(k)\ \tilde{e}^T(k)]^T$. Now, let the signal $v(k)$ be a max-plus-

nonnegative-scaling function of $\tilde{u}(k)$ and $\tilde{e}(k)$ (Van den Boom and De Schutter, 2004):

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

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

$$v(k) = \max_{j=1,\ldots,n_v} (\alpha_j(k) + \beta_j^T \tilde{u}(k) + \gamma_j^T \tilde{e}(k)) \tag{4}$$

Accordingly, the expected value of $v(k)$, denoted by $\hat{v}_{\tilde{u}}(k) = \mathbb{E}[v(k)]$, can be computed analytically as follows:

$$\hat{v}_{\tilde{u}}(k) = \sum_{j=1}^{n_v} \int_{\tilde{e} \in \Phi_j(\tilde{u}(k))} \cdots \int (\alpha_j(k) + \beta_j^T \tilde{u}(k) + \gamma_j^T \tilde{e}) f(\tilde{e}) d\tilde{e} \tag{5}$$

where $d\tilde{e} = d\tilde{e}_1 d\tilde{e}_2 \ldots d\tilde{e}_{n_{\tilde{e}}}$ and the sets $\Phi_j(\tilde{u}(k))$ for $j = 1,\ldots,n_v$ have non-overlapping interiors and are defined such that for all $\tilde{e} \in \Phi_j(\tilde{u}(k))$, $v(k)$ is equal to (4) and $\bigcup_{j=1}^{n_v} \Phi_j(\tilde{u}(k)) = \mathbb{R}^{n_{\tilde{e}}}$. By Proposition 3 of (Van den Boom and De Schutter, 2004) the function $\hat{v}_{\tilde{u}}(k)$ is convex in $\tilde{u}(k)$ and its subgradient $g_v(\tilde{u}(k))$ is given by

$$g_v(\tilde{u}(k)) = \sum_{j=1}^{n_v} \beta_j \int_{\tilde{e} \in \Phi_j(\tilde{u}(k))} \cdots \int f(\tilde{e}) d\tilde{e}. \tag{6}$$

Note that since the system matrices are perturbed by $\tilde{e}(k)$, both $\tilde{y}(k)$ and $\tilde{\kappa}(k)$ depend on $\tilde{e}(k)$. Furthermore, $\mathbb{E}[\tilde{y}(k)]$ and $\mathbb{E}[\tilde{\kappa}(k)]$ are convex in $\tilde{u}(k)$, due to Lemma 3 of (Van den Boom and De Schutter, 2004), which implies that $J_{\text{out}}(k)$ and $J(k)$ are convex in $\tilde{u}(k)$.

Remark 1. It is assumed that the reference signal, $\tilde{r}(k)$, is fixed and known at event step k. The state, $x(k)$, depends on its previous value $x(k-1)$, which is known at event step k, and on the optimal input $\tilde{u}(k)$. Hence, $x(k)$ changes due to the change of $\tilde{u}(k)$. Therefore, we only consider the variations of $\tilde{u}(k)$ and accordingly, the convexity of all functions in $\tilde{u}(k)$.

Hence, referring to Property 4 of (Van den Boom and De Schutter, 2004), if the linear output constraints are monotonically nondecreasing, the MPL-MPC problem turns out to be a convex problem in $\tilde{u}(k)$. Such a problem can be solved using reliable and efficient optimization algorithms, such as interior point methods.

4. APPROXIMATION METHOD

Due to the numerical integration of (5) and (6), the complexity of the direct computation of $\mathbb{E}[\tilde{y}(k)]$ and $\mathbb{E}[\tilde{\kappa}(k)]$ grows fast when the number of stochastic variables, $n_{\tilde{e}}$, increases. To solve this problem, Van den Boom et al. (2007) have introduced an approximation method based on the variability expansion. However, this method could not reduce the complexity of the problem sufficiently and consequently the problem still remains complex. In this section, we introduce an alternative approximation method that is based on the p-th raw moment of a random variable. Related work is presented in (Krivulin, 2000).

We are inspired by considering p-norms and related inequalities (Golub and Van Loan, 1990). Assume that $y = [y_1,\ldots,y_n]^T$ is a vector in $\mathbb{R}^n$; then $\|y\|_p = \left(|y_1|^p + \cdots + |y_n|^p\right)^{1/p}$ and $\|y\|_\infty = \max(|y_1|,\ldots,|y_n|)$ define the p-norm and ∞-norm of y, respectively. However, for the computation of $v(k)$ in (4), we do not consider the absolute value of the variables but the exact values. Obviously, $y \le |y|$ where the equality holds for $y \ge 0$. Hence, to be able to take advantage of the norm relations, we introduce a new variable L as a finite lower bound of y_j, i.e., $L \le y_j$ for $j = 1,\ldots,n$. Hence, $y_j - L \ge 0$, $\forall j$ and we have

$$\begin{aligned}\max(y_1,\ldots,y_n) &= \max(y_1 - L,\ldots,y_n - L) + L \\ &= \max(|y_1 - L|,\ldots,|y_n - L|) + L\end{aligned}$$

Note that for the case that y_j, $j = 1,\ldots,n$, is not bounded from below, we can define L as an offset and consequently, the equality sign changes to inequality as follows:

$$\begin{aligned}\max(y_1,\ldots,y_n) &= \max(y_1 - L,\ldots,y_n - L) + L \\ &\le \max(|y_1 - L|,\ldots,|y_n - L|) + L. \end{aligned} \tag{7}$$

In this case, the role of L is to decrease the error of approximating $y_j - L$ by $|y_j - L|$.

In the sequel we will use the following theorem (Boyd and Vandenberghe, 2004):

Theorem 2. (Jensen's Inequality). Let x be an integrable real-valued random variable and φ be an integrable, concave function. Then: $\varphi(\mathbb{E}[x]) \ge \mathbb{E}[\varphi(x)]$.

In this paper, we assume that each element of the error vector $\tilde{e}_i(k)$ is an i.i.d. normally distributed random variable [1] with mean μ_i and variance σ_i^2, i.e., $\tilde{e}_i(k) \sim \mathcal{N}(\mu_i, \sigma_i^2)$ for $i = 1,\ldots,n_{\tilde{e}}$. Hence, $\tilde{e}_i(k)$ is not bounded from below. However, 99.7% of the observations of a normally distributed random variable fall within 3 standard deviation of the mean, i.e., between $\mu_i - 3\sigma_i$ and $\mu_i + 3\sigma_i$. Therefore, the lower bound of each random variable $\tilde{e}_i(k)$ can be approximated by $\zeta_i = \mu_i - 3\sigma_i$ for $i = 1,\ldots,n_{\tilde{e}}$ and hence, we can take $L = \min(\zeta_1,\ldots,\zeta_{n_{\tilde{e}}})$.

Now consider the random variables $y_j \sim \mathcal{N}(\mu_j, \sigma_j^2)$ for $j = 1,\ldots,n$. Let $x_j = y_j - L$ where $L = \min_{j=1,\ldots,n} (\mu_j - 3\sigma_j)$. By considering (7) and the fact that expected value is a linear, monotonic operator, we have:

$$\begin{aligned}\mathbb{E}\big[\max(x_1,\ldots,x_n)\big] + L &\le \mathbb{E}\big[\max(|x_1|,\ldots,|x_n|)\big] + L \\ &\le \mathbb{E}\|x\|_\infty + L \end{aligned} \tag{8}$$

Since $\|x\|_\infty \le \|x\|_p$ (Golub and Van Loan, 1990), we have:

$$\begin{aligned}&\mathbb{E}\big[\|x\|_\infty + L\big] \le \mathbb{E}\big[(|x_1|^p + \cdots + |x_n|^p)^{1/p} + L\big] \\ &\text{or } \mathbb{E}\|x\|_\infty + L \le \mathbb{E}\big[(|x_1|^p + \cdots + |x_n|^p)^{1/p}\big] + L \end{aligned} \tag{9}$$

Finally, by applying Theorem 2 to the right-hand side of (9), we obtain:

$$\mathbb{E}\|x\|_\infty + L \le \left(\mathbb{E}\big[|x_1|^p + \cdots + |x_n|^p\big]\right)^{1/p} + L \tag{10}$$

Note that we can apply Theorem 2 since $\varphi(x) = x^{1/p}$ is a concave function for $p > 1$ and $x > 0$, and in our case the argument x is $\sum_{i=1}^n |x_i|^p$ which is positive. Accordingly, from (8)-(10) we conclude that

$$\mathbb{E}\big[\max(x_1,\ldots,x_n)\big] + L \le \left(\sum_{j=1}^{n} \mathbb{E}\big[|x_j|^p\big]\right)^{1/p} + L \tag{11}$$

Therefore, instead of minimizing the left-hand side of (11), we can minimize its right-hand side. As a result, we choose

[1] Note that in theory this may sometimes result in negative values for quantities that are in principle positive such as the execution times. However, this will not have any impact later on due to the selection of the offset or lower bound L (which will then be taken positive).

the function $\hat{v}_{\tilde{u},\text{app}}(k)$ as an approximation of $\hat{v}_{\tilde{u}}(k)$ for an appropriate choice of p as follows [2] :

$$\hat{v}_{\tilde{u},\text{app}}(k) = \left(\sum_{j=1}^{n_v} \mathbb{E}\big[(\alpha_j + \beta_j^T \tilde{u} + \gamma_j^T \tilde{e} - L)^p \big] \right)^{1/p} + L. \quad (12)$$

Remark 3. For an even positive integer $p = 2q$, $q \in \mathbb{Z}^+$, $\mathbb{E}[x^p] = \mathbb{E}[|x|^p]$. Hence, without loss of generality, we can use $\mathbb{E}[x^p]$ in our problem by only considering even values for p in the sequel. So from now on, p is an even integer larger than or equal to 2.

We also have (Dekking et al., 2005):

Lemma 4. If $z_1 \sim \mathcal{N}(\mu_1, \sigma_1^2)$ and $z_2 \sim \mathcal{N}(\mu_2, \sigma_2^2)$ are independent, then $az_1 + bz_2 + c \sim \mathcal{N}(a\mu_1 + b\mu_2 + c, a^2\sigma_1^2 + b^2\sigma_2^2)$ for any real numbers a, b, c

Recall that we assume that each element of the error vector is normally distributed, i.e., $\tilde{e}_i \sim \mathcal{N}(\mu_i, \sigma_i^2)$ for $i = 1, \ldots, n_{\tilde{e}}$. By Lemma 4, the random variable $x_j = \alpha_j + \beta_j^T \tilde{u} + \gamma_j^T \tilde{e} - L$ in (12) is also normally distributed with mean $\mu_j = \alpha_j + \beta_j^T \tilde{u} + \gamma_j^T \mu_{\tilde{e}} - L$ and variance $\sigma_j^2 = \gamma_j^T [\sigma_1^2, \ldots, \sigma_{n_{\tilde{e}}}^2]^T$. Note that $L = \min_{j=1,\ldots,n_v} (\alpha_j + \beta_j^T \tilde{u} + \gamma_j^T \mu_{\tilde{e}} - 3\sigma_j)$ and $\mu_{\tilde{e}} = [\mu_1, \ldots, \mu_{n_{\tilde{e}}}]^T$ is the mean of the error vector $\tilde{e}$.

By definition, the p-th raw moment of a normally distributed random variable x with mean μ and standard deviation σ can be computed as follows:

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

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

$$\mathbb{E}\big[x^p\big] = \sigma^p i^{-p} H_p(i\mu/\sigma) \quad (13)$$

where

$$H_p(x) \equiv (-1)^p \exp(x^2/2) \frac{d^p}{dx^p} \exp(-x^2/2)$$

is the p-th Hermite polynomial. Considering equations (26.2.51) and (22.3.11) in (Abramowitz and Stegun, 1964), leads to

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

where p is an even integer in our case and therefore, $p/2 \in \mathbb{Z}^+$. As a consequence, $\mathbb{E}\big[x^p\big]$ in (13) can be written as,

$$\mathbb{E}\big[x^p\big] = \sum_{k=0}^{p/2} \sigma^{2k} 2^{-k} \frac{p!}{k!(p-2k)!} \mu^{p-2k}, \quad (14)$$

which gives a better insight into the construction of (13).

5. CONVEXITY OF THE APPROXIMATION

The approximation method (12), where the expectation can be replaced by (13) or equivalently (14), decreases the computational complexity significantly since there is no numerical integration involved any more and consequently, increases the time efficiency. Moreover, if we can prove the convexity of (12), it is possible to use convex optimization algorithms to solve the MPC optimization problem in a very efficient way.

To this end, let $f_j(\tilde{u}) = \alpha_j + \beta_j^T \tilde{u} + \gamma_j^T \tilde{e} - L$, which is an affine and so a convex function in $\tilde{u}$. As mentioned in Section 4, the p-th moment of a normally distributed random variable is finite for $p \geq 2$. Consequently, since $\tilde{e}$ is normally distributed, $\mathbb{E}[(\tilde{e}(k))^p] < \infty$. Accordingly, the p-th raw moment of $f_j(\tilde{u})$ is defined as:

$$\mathbb{E}[f_j(\tilde{u})^p] = \int_{-\infty}^{\infty} \Big(f_j(u) \Big)^p \, \mathrm{d}F_{\tilde{u}}(u)$$

where $F_{\tilde{u}}$ is the cumulative distribution function of $f_j(\tilde{u})$. Note that $f_j(\tilde{u})$ is normally distributed with mean $\mu_j = \alpha_j + \beta_j^T \tilde{u}(k) + \gamma_j^T \mu_{\tilde{e}} - L$ and variance $\sigma_j^2 = \gamma_j^T [\sigma_1^2, \ldots, \sigma_{n_{\tilde{e}}}^2]^T$, as mentioned in Section 4. To be able to proceed further, we need the following theorems (see Chapter 5 of (Mitrinović et al., 1993)):

Theorem 5. (Minkowski inequality for functions). Let h and g be real-valued functions in $\mathbb{R}$ such that the functions $|h(x)|^\ell$ and $|g(x)|^\ell$ for an integer $\ell > 1$ are integrable on $\mathbb{R}$. Then

$$\Big(\mathbb{E}\big[|h(x) + g(x)|^\ell \big] \Big)^{1/\ell} \leq \Big(\mathbb{E}\big[|h(x)|^\ell \big] \Big)^{1/\ell} + \Big(\mathbb{E}\big[|g(x)|^\ell \big] \Big)^{1/\ell}$$

Theorem 6. (Minkowski inequality for vectors). Let $x, y \in \mathbb{R}^n$ and $\ell > 1$ be an integer. Then,

$$\left(\sum_{j=1}^{n} |x_j + y_j|^\ell \right)^{1/\ell} \leq \left(\sum_{j=1}^{n} |x_j|^\ell \right)^{1/\ell} + \left(\sum_{j=1}^{n} |y_j|^\ell \right)^{1/\ell}.$$

Since p is even in our case (cf. Remark 3), we have $|x|^p = x^p$. Consequently, we do not use the absolute value sign for the expressions with the power p in the rest of this section. First we prove the following proposition:

Proposition 7. $\big(\mathbb{E}\big[\big(f_j(\tilde{u}) \big)^p \big] \big)^{1/p}$ is a convex function of $\tilde{u}$.

Proof: If we show that

$$\Big(\mathbb{E}\big[\big(f_j(\lambda \tilde{u}_1 + (1-\lambda)\tilde{u}_2) \big)^p \big] \Big)^{1/p}$$
$$\leq \lambda \Big(\mathbb{E}[f_j(\tilde{u}_1)^p] \Big)^{1/p} + (1-\lambda) \Big(\mathbb{E}[f_j(\tilde{u}_2)^p] \Big)^{1/p}$$

for any two points $\tilde{u}_1$ and $\tilde{u}_2$ in the domain of $\mathbb{E}\big[\big(f_j(\tilde{u}) \big)^p \big]$ and for any $0 \leq \lambda \leq 1$, then the proof is complete. Since f_j is a linear function in $\tilde{u}$, we have

$$f_j(\lambda \tilde{u}_1 + (1-\lambda)\tilde{u}_2) = \lambda f_j(\tilde{u}_1) + (1-\lambda) f_j(\tilde{u}_2)$$

Therefore, instead of the inequality above, we prove that the following inequality holds true.

$$\Big(\mathbb{E}\big[\big(\lambda f_j(\tilde{u}_1) + (1-\lambda) f_j(\tilde{u}_2) \big)^p \big] \Big)^{1/p}$$
$$\leq \lambda \Big(\mathbb{E}[f_j(\tilde{u}_1)^p] \Big)^{1/p} + (1-\lambda) \Big(\mathbb{E}[f_j(\tilde{u}_2)^p] \Big)^{1/p} \quad (15)$$

In fact (15) follows directly from the Minkowski inequality for functions (Theorem 5). So the inequality holds true and consequently, $\big(\mathbb{E}\big[\big(f_j(\tilde{u}) \big)^p \big] \big)^{1/p}$ is a convex function in $\tilde{u}$. □

Before proving the convexity of (12), consider the following remark.

Remark 8. Let $x, y \in \mathbb{R}^n$ and $p \geq 2$ be an even integer. If $|x_j| < |y_j|$ for $j = 1, \ldots, n$, then

$$\left(\sum_{j=1}^{n} |x_j|^p \right)^{1/p} < \left(\sum_{j=1}^{n} |y_j|^p \right)^{1/p}.$$

Now according to the above notation, we can rewrite (12) more compactly as

$$\hat{v}_{\tilde{u},\text{app}}(k) = \Big(\sum_{j=1}^{n_v} \mathbb{E}[f_j(\tilde{u})^p] \Big)^{1/p} + L.$$

[2] For brevity, we drop k everywhere, except for $\hat{v}$.

Proposition 9. $\hat{v}_{\tilde{u},\mathrm{app}}(k)$ is a convex function of $\tilde{u}$.

Proof: Considering Remark 8, the Minkowski inequality for vectors (Theorem 6), and Proposition 7, we prove that $\hat{v}_{\lambda\tilde{u}_1+(1-\lambda)\tilde{u}_2,\mathrm{app}}(k) \leq \lambda\hat{v}_{\tilde{u}_1,\mathrm{app}}(k) + (1-\lambda)\hat{v}_{\tilde{u}_2,\mathrm{app}}(k)$ and therefore, it is a convex function of $\tilde{u}$:

$$
\begin{aligned}
&\left(\sum_{j=1}^{n_v}\mathbb{E}\left[\left(\lambda f_j(\tilde{u}_1)+(1-\lambda)f_j(\tilde{u}_2)\right)^p\right]\right)^{1/p}\\
&=\left(\sum_{j=1}^{n_v}\left(\mathbb{E}\left[\left(\lambda f_j(\tilde{u}_1)+(1-\lambda)f_j(\tilde{u}_2)\right)^p\right]\right)^{(1/p)(p)}\right)^{1/p}\\
&\overset{\text{eq.(15),Rem.8}}{\leq}\left(\sum_{j=1}^{n_v}\Big(\lambda\left(\mathbb{E}[f_j(\tilde{u}_1)^p]\right)^{1/p}\right.\\
&\qquad\left.+(1-\lambda)\left(\mathbb{E}[f_j(\tilde{u}_2)^p]\right)^{1/p}\Big)^p\right)^{1/p}\\
&\overset{\text{Minkowski, Thm.6}}{\leq}\left(\sum_{j=1}^{n_v}\lambda^p\mathbb{E}[f_j(\tilde{u}_1)^p]\right)^{1/p}\\
&\qquad+\left(\sum_{j=1}^{n_v}(1-\lambda)^p\mathbb{E}[f_j(\tilde{u}_2)^p]\right)^{1/p}\\
&\leq\lambda\left(\sum_{j=1}^{n_v}\mathbb{E}[f_j(\tilde{u}_1)^p]\right)^{1/p}+(1-\lambda)\left(\sum_{j=1}^{n_v}\mathbb{E}[f_j(\tilde{u}_2)^p]\right)^{1/p}
\end{aligned}
$$

Note that the constant value L in $\hat{v}_{\tilde{u},\mathrm{app}}(k)$ is omitted since this term does not influence the convexity. □

6. WORKED EXAMPLE

Now we illustrate our approximation method using an example that involves a simple manufacturing system taken from Van den Boom and De Schutter (2004). Selecting this example also enables us to compare our results with the exact analytic solution. Moreover, we can also compare the computation time of our method with the one of (Van den Boom and De Schutter, 2004).

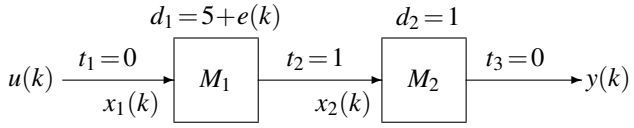

Fig. 1. A production system.

Consider the manufacturing system of Figure 1. This system consists of two machines M_1 and M_2 and operates in batches. The raw material is fed to M_1 where preprocessing is done. Afterwards the intermediate product is fed to M_2 and finally leaves the system. We assume that each machine starts working as soon as possible on each batch, i.e., as soon as the raw material or the required intermediate product is available, and as soon as the machine is idle (i.e., the previous batch of products has been processed and has left the machine).

Define:

$u(k)$: time instant at which the system is fed 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 Van den Boom and De Schutter (2004)

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

In matrix notation this becomes

$$
\begin{aligned}
x(k) &= A(k)\otimes x(k-1)\oplus B(k)\otimes u(k)\\
y(k) &= C(k)\otimes x(k)\ .
\end{aligned}
$$

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

$$
\begin{aligned}
A &= \begin{bmatrix} d_1(k-1) & \varepsilon \\ d_1(k-1)+d_1(k)+t_2(k) & d_2(k-1)\end{bmatrix}\\
B &= \begin{bmatrix} t_1(k) \\ d_1(k)+t_1(k)+t_2(k)\end{bmatrix}\\
C &= \begin{bmatrix}\varepsilon & d_2(k)+t_3(k)\end{bmatrix}.
\end{aligned}
$$

It is assumed that the transportation times are constant: $t_1(k) = 0, t_2(k) = 1, t_3(k) = 0$; the production time of M_2 is constant: $d_2(k) = 1$; the due date (reference) signal is $r(k) = 4+6\cdot k$ and the initial state is $x(0) = [0\ \ 7]^T$. The vector $\tilde{e}(k) = [\,d_1(k-1)\ \ d_1(k)\ \ d_1(k+1)\ \ d_1(k+2)\,]^T$ consists of independent stochastic random variables, i.e., the production time on M_1 is corrupted by noise: $d_1(k+\ell) = 5+e(k+\ell)$ where $e(k+\ell) \sim \mathcal{N}(0,1)$ for $\ell = -1,\ldots,2$. The cost criterion (3) will be optimized for $N_\mathrm{p} = 3$ and $N_\mathrm{c} = 2$. Hence, we can write the tardiness error equations with deterministic and stochastic parts:

$$
\begin{aligned}
&\max(y(k)-r(k),0)\\
&\quad=\max(\eta_1+\tilde{e}_1+\tilde{e}_2, \eta_2+\tilde{e}_2, \eta_3, 0)\\
&\max(y(k+1)-r(k+1),0)\\
&\quad=\max(\eta_4+\tilde{e}_1+\tilde{e}_2+\tilde{e}_3, \eta_5+\tilde{e}_2+\tilde{e}_3,\\
&\qquad\eta_6+\tilde{e}_3, \eta_7+\tilde{e}_1+\tilde{e}_2, \eta_8+\tilde{e}_2, \eta_9, 0)\\
&\max(y(k+2)-r(k+2),0)\\
&\quad=\max(\eta_{10}+\tilde{e}_1+\tilde{e}_2+\tilde{e}_3+\tilde{e}_4, \eta_{11}+\tilde{e}_2+\tilde{e}_3+\tilde{e}_4,\\
&\qquad\eta_{12}+\tilde{e}_3+\tilde{e}_4, \eta_{13}+\tilde{e}_4, \eta_{14}+\tilde{e}_1+\tilde{e}_2+\tilde{e}_3,\\
&\qquad\eta_{15}+\tilde{e}_2+\tilde{e}_3, \eta_{16}+\tilde{e}_3, \eta_{17}+\tilde{e}_1+\tilde{e}_2,\\
&\qquad\eta_{18}+\tilde{e}_2, \eta_{19}, 0)
\end{aligned}
\tag{16}
$$

where $\eta_1,\ldots,\eta_{19}$ are sums of deterministic values and $\tilde{e}_1,\ldots,\tilde{e}_4$ are the elements of the stochastic vector $\tilde{e}(k)$. Note that we omit the argument k for brevity. Next, we add an offset L to all terms in $y(k+\ell)-r(k+\ell), \ell = 0,1,2$. As mentioned in Section 4 the value L used in (7) for each maximization term will be defined as follows:

$$
L = \min_{j=1,\ldots,n_v}(\eta_j+\gamma_j^T\mu_{\tilde{e}}-3\sigma_j, 0)
$$

where $\mu_{\tilde{e}} = [\mu_1,\ldots,\mu_4]^T$ and $\sigma_j^2 = \gamma_j^T[\sigma_1^2,\ldots,\sigma_4^2]^T$ such that $\mu_i = 0$ and $\sigma_i^2 = 1$ are the mean and variance of $\tilde{e}_i$ for $i = 1,\ldots,4$ by assumption, and γ_j is a vector of 0 and 1 that indicates which elements of $\tilde{e}$ appear in the maximization.

Table 1 lists the results of computing the expected values of the tardiness error shown in (16) using the approximation method

k	ℓ	$\hat{v}_{\tilde{u}}(k)$	$\hat{v}_{\tilde{u},\mathrm{app}}(k)$	relative error (%)
1	0	2.0089	3.6895	84
	1	6.0120	6.2912	4.644
	2	11.0099	11.1758	1.507
2	0	3.5606	3.6318	2.003
	1	8.5607	8.5637	0.035
	2	13.5603	13.5638	0.026
3	0	5.8460	5.8476	0.027
	1	10.8460	10.8464	0.004
	2	15.8459	15.8465	0.004
4	0	4.9771	5.0006	0.472
	1	9.9766	10.0011	0.246
	2	14.9758	15.0036	0.186

Table 1. Exact value $\hat{v}_{\tilde{u}}(k)$ and approximation $\hat{v}_{\tilde{u},\mathrm{app}}(k)$ of $\max(y(k+\ell)-r(k+\ell),0)$ for $\ell = 0,1,2$.

(12), as well as the exact values obtained by using numerical integration. At each event step $k = 1,2,3,4$, three expectations are computed due to the equations in (16). Our numerical experiment with different values of p show that for the first expectation, $p = 8$, for the second one, $p = 24$, and for the third one, $p = 36$, give a very good approximation. The error of the approximation method is less than 2% on average, as shown in the last column of Table 1 which presents the percentage relative error.

If we apply MPC in closed loop for 20 successive event steps, then the total computation time [3] using our approximation approach is 9 s. The computation time for the method of Van den Boom and De Schutter (2004) when applied to this example is 93 s, but this computation time holds for the case that the error vector $\tilde{e}$ is *uniformly* distributed; for the *normal* distribution the computation time will only increase.

So the proposed approximation approach combines a good performance with computation times that are much lower than existing approaches, and is thus an efficient and useful tool for solving the MPL-MPC optimization problem.

7. CONCLUSIONS

For the stochastic MPL-MPC problem, we have introduced an approximation method to reduce the computational complexity. In this stochastic framework, the complexity of the optimization problem increases if the number of stochastic variables increases, due to the numerical integrations present in the cost function and output constraint. To tackle these difficulties, we have proposed a method based on the p-th raw moments of the stochastic variables. We assumed that the elements of the noise vector in the system are normally distributed. Since a closed form of the p-th raw moments of a normally distributed random variable exists, we can compute the expectation without using the numerical integration. We have also proved that the expectations and consequently the cost function are convex. Therefore, if the constraints are a nondecreasing affine function of the output, the resulting optimization problem is a convex optimization problem, and thus can be solved very efficiently.

Future work will focus on decreasing the error of approximation method by finding the best choice for the offset L and the raw moment order p. Moreover, we will compare this method with variability expansion (Van den Boom et al., 2007) more precisely to find the strong and weak points of these methods in comparison to each other.

[3] The CPU times reported here were obtained running Matlab 7.5.0 (R2007b) on a 2.33 GHz Intel Core Duo E655 processor.

REFERENCES

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

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

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

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

De Schutter, B. and van den Boom, T. (2001). Model predictive control for max-plus-linear discrete event systems. *Automatica*, 37(7), 1049–1056.

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

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

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

Krivulin, N. (2000). Algebraic modeling and performance evaluation of acyclic fork-join queueing networks. *Advances in Stochastic Simulation Methods*, 63–82.

Mitrinović, D.S., Pečrić, J.E., and Fink, A.M. (1993). *Classical and New Inequalities in Analysis*. Kluwer Academic Publishers, Dordrecht, Netherlands.

Necoara, I., De Schutter, B., van den Boom, T., and Hellendoorn, J. (2004). Model predictive control for perturbed continuous piecewise affine systems with bounded disturbances. In *Proceedings of the 43rd IEEE Conference on Decision and Control*, 1848–1853. Paradise Island, The Bahamas.

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

Rawlings, J. and Mayne, D. (2009). *Model Predictive Control: Theory and Design*. Nob Hill Publishing, Madison, Wisconsin.

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

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

Van den Boom, T.J.J., Heidergott, B., and De Schutter, B. (2007). Complexity reduction in MPC for stochastic max-plus-linear discrete event systems by variability expansion. *Automatica*, 43(6), 1058–1063.

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