Fast Stochastic Model Predictive Control of High-dimensional Systems

Joel A. Paulson¹, Ali Mesbah¹, Stefan Streif², Rolf Findeisen², and Richard D. Braatz¹

Abstract—Probabilistic uncertainties and constraints are ubiquitous in complex dynamical systems and can lead to severe closed-loop performance degradation. This paper presents a fast algorithm for stochastic model predictive control (SMPC) of high-dimensional stable linear systems with time-invariant probabilistic uncertainties in initial conditions and system parameters. Tools and concepts from polynomial chaos theory and quadratic dynamic matrix control inform the development of an input-output formulation for SMPC with output constraints. Generalized polynomial chaos theory is used to enable efficient uncertainty propagation through the high-dimensional system model. Galerkin projection is used to construct the polynomial chaos expansion for a general class of linear differential algebraic equations (DAEs), so that the SMPC algorithm is applicable to both regular and singular/descriptor systems. The fast SMPC approach is applied for control of an end-to-end continuous pharmaceutical manufacturing process with approximately 8000 states. The on-line computational cost of the proposed probabilistic input-output SMPC algorithm is independent of the state dimension and, therefore, alleviates the prohibitive computational costs of control of uncertain systems with large state dimension.

I. INTRODUCTION

Model predictive control (MPC) is the most widely used approach for the advanced control of complex dynamical systems due to its ability to systematically deal with multivariable dynamics, system constraints, and competing sets of objectives [1]. While many MPC formulations have been developed to account for exogenous disturbances and measurement noise, few formulations are able to systematically handle probabilistic uncertain parameters, in spite of their ubiquity in complex systems. Probabilistic uncertainties can lead to severe closed-loop performance degradation and, as a result, impair high-performance operation of complex systems using classical MPC approaches.

Robust MPC is a popular approach for dealing with model uncertainties (e.g., see [2]). Assuming that uncertainties are bounded, most robust MPC approaches compute the optimal control law to minimize the performance for the worst-case model uncertainties. While this approach can guarantee constraint satisfaction under all disturbances and uncertainties, the resulting control law is conservative in most control applications, as the worst-case perturbations usually have a very low probability of occurrence. Recently, novel methods for robust MPC have been investigated using the so-called scenario approach (e.g., see [3], [4], and the references therein). The scenario approach provides a sampling-based technique to solve convex chance-constrained optimization problems and, therefore, enables a paradigm shift from deterministic algorithms to randomized robust MPC approaches that exploit the statistical description of uncertainties. However, typically the number of samples (i.e., scenarios) results in an on-line computational cost that is too high for implementation in high-dimensional systems.

Stochastic MPC (SMPC) offers an alternative approach for robust MPC in a probabilistic uncertainty setting. SMPC approaches enable shaping the predicted probability distribution functions (PDFs) of system states and outputs in an optimal manner over a finite prediction horizon (e.g., see [5], [6], [7], [8], [9], and the references therein). In SMPC, chance constraints can be considered in a probabilistic sense to circumvent the inherent conservatism of deterministic worst-case robust MPC approaches. However, a key challenge in SMPC is the propagation of probabilistic uncertainties through the system model. The commonly used approaches for probabilistic uncertainty analysis (e.g., Monte Carlo methods [10]) are prohibitively expensive for real-time control for high-dimensional systems.

This work presents a fast stochastic MPC algorithm applicable to high-dimensional stable systems with time-invariant probabilistic uncertainties in initial conditions and system parameters. The quadratic dynamic matrix control (QDMC) algorithm [11], which is the MPC algorithm most widely applied to large-scale industrial systems, is adopted to formulate an input-output framework for SMPC with output constraints. Such a probabilistic input-output framework has an online computational cost that is independent of the state dimension, which enables its application to uncertain systems with high state dimension.¹

Generalized polynomial chaos theory [12], [13] is used for propagation of probabilistic uncertainties through the high-dimensional system model. In polynomial chaos theory, the implicit mappings between uncertain variables/parameters and the system states are replaced with expansions of orthogonal polynomials that are functions of the random variables. In contrast to sampling-based uncertainty analysis approaches, the orthogonality property of polynomial chaos expansions (PCEs) enables efficient computation of the statistical properties of the state PDFs. This paper uses Galerkin projection to determine the coefficients of the PCEs for a general class of linear differential algebraic equations (DAEs). This paper generalizes the application of

¹Such systems are referred to as high-dimensional systems in this paper.
the proposed fast SMPC approach to a large class of complex systems, whose dynamics are described by a set of high-dimensional DAEs.

The next section formulates the stochastic MPC problem for high-dimensional systems with probabilistic uncertainties. Sec. III presents polynomial chaos expansions for uncertainty propagation, where the Galerkin projection method is introduced to linear dynamical systems described by DAEs. The quadratic dynamic matrix control algorithm is introduced in Sec. IV to formulate the MPC problem in terms of system inputs and outputs, with on-line computational cost independent of the state dimension. Sec. V presents the proposed fast SMPC approach, which is applied to the control of a high-dimensional continuous pharmaceutical manufacturing process in the presence of parametric uncertainties (Sec. VI).

**II. PROBLEM STATEMENT**

Consider an uncertain continuous-time stable linear differential-algebraic system,

\[
M(\theta)\dot{x}(t, \theta) = A(\theta)x(t, \theta) + B(\theta)u(t) + r(\theta)
\]

\[x(0, \theta) = x_0(\theta)\]  

(1)

where \(x \in \mathbb{R}^{n_x}\) denotes the system states, \(\dot{x}\) denotes the derivative of \(x\) with respect to time \(t\), \(x_0\) denotes the initial conditions, \(u \in \mathbb{R}^{n_u}\) denotes the inputs to the system, and \(\theta \in \mathbb{R}^{n_\theta}\) denotes the system random variables (i.e., uncertain parameters and initial conditions). Eq. (1) is a differential algebraic equation (DAE) and is assumed to be in its equivalent index-1 form (e.g., see [14]). The state vector \(x = [x^T_x, x^T_a]^T\) is composed of differential states \(x_d \in \mathbb{R}^{n_d}\) (i.e., states whose derivatives appear in the vector \(M(\theta)\dot{x}(t, \theta)\) in (1)) and algebraic states \(x_a \in \mathbb{R}^{n_a}\) (i.e., the rest of the states) with \(n_x = n_d + n_a\). For a consistent initial condition in (1), \(n_x + n_d\) variables must be specified \((x(0, \theta)\) and \(\dot{x}_d(0, \theta))\), with \(n_d\) degrees of freedom (DOF) set by (1) holding at \(t = 0\) and \(n_d\) additional DOF. The system outputs, denoted by \(y \in \mathbb{R}^{n_y}\), are algebraically related to \(x, u, \) and \(\theta\). Hence, \(y\) can also be treated as algebraic states and can be included in the definition of \(x\) in (1) for notational simplicity. The vector \(r \in \mathbb{R}^{n_r}\) is used to represent time-invariant stochastic behavior such as process disturbances and actuator response variations due to, for example, static friction and backlash. Note that \(r\) can also be used to represent additional terms or error from linearization of nonlinear dynamics.

The system (1) is often called a descriptor or singular system in the control literature (e.g., [15], [16] and references cited therein), with systems described by ordinary differential equations being a special case. The extra algebraic terms enable the representation of a much broader class of systems, including pH neutralization, electrochemical systems, and some classes of electronic and mechanical systems.

In (1), \(\theta\) is composed of independently distributed random variables \(\theta_i\) with known probability distribution functions (PDFs) \(f_{\theta_i}\). A probability triple \((\Omega, \mathcal{F}, P)\) is defined on the basis of sample space \(\Omega\), \(\sigma\)-algebra \(\mathcal{F}\), and probability measure \(P\) on \((\Omega, \mathcal{F})\). The PDFs are defined such that \(\theta_i \in \mathcal{L}^2(\Omega, \mathcal{F}, P), \forall i \in \{1, \ldots, n_\theta\}\), where \(\mathcal{L}^2(\Omega, \mathcal{F}, P)\) represents the Hilbert space of all random variables \(\theta_i\) with finite \(L_2\) norm. The expected value (first-order moment) of a stochastic variable \(\psi : \Omega \rightarrow \mathbb{R}\) is denoted by \(E[\psi] := \int_\Omega \psi d\psi\), where \(d\psi\) is the PDF of \(\psi\) over its support \(\Omega\). The variance (central second-order moment) of \(\psi\) is denoted by \(\text{Var}[\psi] := E[(\psi - E[\psi])^2]\).

Due to the probabilistic uncertainties \(\theta\), the solution trajectories of system (1) are probabilistically distributed. In this work, the goal of controller synthesis is to shape the probability distributions of system outputs to have desirable statistics. Assuming that the system states can be estimated at all times, a finite-horizon SMPC problem can be stated as follows.

**Problem 1. (Finite-horizon stochastic MPC):**

\[
\min_{u(t)} \quad J(\tilde{x}_d(t_k), u(t)) \quad \text{(2)}
\]

s.t.:  
\[M\tilde{x}(t, \theta) = A\tilde{x}(t, \theta) + Bu(t) + r, \quad t_k \leq t \leq t_p\]
\[A_h\tilde{y}(t, \theta) \leq b_h, \quad t_k \leq t \leq t_p\]
\[u(t) \in \mathcal{U}, \quad t_k \leq t \leq t_m\]
\[\tilde{x}_d(0, \theta) = \tilde{x}_d(t_k)\]

where \(u(t), t \in [t_k, t_m]\), denotes the input profile (control policy), \(t_m\) denotes the control horizon, \(t_p\) denotes the prediction horizon, \(\tilde{x}(t, \theta)\) denotes the predicted states from the DAE system model in (1), \(\tilde{y}(t, \theta)\) denotes the predicted outputs (which are also lumped into the vector \(\tilde{x}\)), \(\tilde{x}_d(k)\) denotes the estimated differential states at time instant \(t_k\) computed from measurements, \((A_h, b_h)\) specifies the linear constraints on the outputs, and \(\mathcal{U} \subset \mathbb{R}^{n_u}\) denotes the convex compact set of input constraints. Problem 1 specifies an optimal input profile based on the uncertain DAE model (1) and the estimated state variables \(\tilde{x}_d(t_k)\), while satisfying hard constraints on the inputs and general inequality constraints on the outputs. The cost function \(J(\tilde{x}_d(k), u(t))\) is commonly defined in terms of some statistics or moments of the PDFs of the outputs (e.g., [9]). Note that \(M, A, B, \) and \(r\) are all functions of the random variables \(\theta\) (dependency on \(\theta\) is not explicitly written to shorten the notation), and a unique consistent initialization of (1) can be computed from \(\tilde{x}_d(0, \theta)\).

Solving the SMPC problem (2) in a receding-horizon mode for high-dimensional systems is particularly challenging. The difficulties arise from (i) the prohibitive computational costs of model simulation and optimization for on-line control due to a large state dimension, (ii) the need for high-dimensional state estimation to determine \(\tilde{x}_d(k)\) from limited system measurements (even more difficult since the system is most likely not observable), and (iii) the propagation of probabilistic uncertainties \(\theta\) through the system model (1). To address challenges (i) and (ii), the QDMC algorithm is used in this work to reformulate the high-dimensional control problem in terms of system inputs and outputs, which are usually of much lower dimensions than the system states in real applications. The input-output framework of QDMC not only eliminates the high computational costs.
associated with on-line control of high-dimensional systems, but also alleviates the need for high-dimensional state estimation since the output measurements can be readily incorporated into the control algorithm to update the system model. To enable efficient uncertainty analysis, polynomial chaos expansions (PCEs) are used to propagate the uncertainties \( \theta \) through the dynamics in (1). PCEs are advantageous as they can be written directly in terms of system inputs and outputs, and enable efficient computation of the statistical moments of outputs. Next, the main principles of PCEs are introduced, including the determination of the PCE coefficients for linear DAE systems by Galerkin projection.

III. UNCERTAINTY PROPAGATION

A. Polynomial Chaos Expansion

Polynomial chaos expansions provide a means for approximating a stochastic variable \( \psi(\theta) \in L^2(\Omega, F, P) \) with the \( L_2 \)-convergent expansion [13]

\[
\psi(\theta) = \sum_{k=0}^{\infty} a_k \Phi_k(\theta)
\]

where \( a_k \) denotes the expansion coefficients and \( \Phi_k(\theta) \) denotes the polynomial chaos basis functions of degree \( m \) with respect to the random variables \( \theta \). These basis functions belong to the Askey scheme of polynomials, which encompasses a set of orthogonal basis functions in the Hilbert space defined by the support of the random variables. Hence, the basis functions satisfy \( \langle \Phi_i(\theta), \Phi_j(\theta) \rangle = \delta_{ij} \) where \( \langle h(\theta), g(\theta) \rangle = \int_{\Omega} h(\theta) g(\theta) d\theta = E[h(\theta)g(\theta)] \) denotes the inner product with respect to the weight \( f_\theta \) (PDF of \( \theta \)) and over the domain \( \Omega \) (support of \( \theta \)). The choice of orthogonal polynomials is made such that their weight function is the multivariate PDF of \( \theta \). Table I shows the orthogonal polynomials corresponding to particular distributions in \( \theta \).

For practical reasons, the PCE in (3) must be truncated to a finite number of terms. The total number of terms \( L + 1 = (n_0 + m_0)^2 \) in the truncated expansion depends on the number of uncertain parameters \( n_0 \) and the highest order of the polynomial basis functions \( m \) retained in the expansion

\[
\hat{\psi}(\theta) := \sum_{k=0}^{L} a_k \Phi_k(\theta) = a^T \Lambda(\theta)
\]

with \( a = [a_0, \ldots, a_L]^T \) and \( \Lambda(\theta) = [\Phi_0(\theta), \ldots, \Phi_L(\theta)]^T \). The vector of PCE coefficients \( a \) can be computed using probabilistic collocation methods or Galerkin projection, depending on the complexity of system dynamics (e.g., see [13], [17], [18] and references therein). The orthogonality property of the multivariate polynomials can be used to efficiently compute the PDF statistics of the stochastic variable \( \psi(\theta) \). For instance, the first- and second-order central moments of \( \psi(\theta) \) are defined by

\[
E[\hat{\psi}(\theta)] = a_0
\]

\[
\text{Var}[\hat{\psi}(\theta)] = \sum_{k=1}^{L} a_k^2 \langle \Phi_k(\theta)^2 \rangle.
\]

These moments, along with higher order moments, can be used to approximate the PDF of \( \psi(\theta) \) [19].

<table>
<thead>
<tr>
<th>POLYNOMIALS ( \Phi_k ) CORRESPONDING TO THE PDF OF ( \theta_i ). THE POLYNOMIALS ARE FUNCTIONS OF THE STANDARD PDF OVER THE LISTED SUPPORT. METHODS EXIST FOR CONVERTING THE SUPPORT OF ARBITRARY PDFS TO AND FROM THE STANDARD FORM [13]</th>
</tr>
</thead>
<tbody>
<tr>
<td>PDF</td>
</tr>
<tr>
<td>---</td>
</tr>
<tr>
<td>Beta</td>
</tr>
<tr>
<td>Chebyshev</td>
</tr>
<tr>
<td>Gamma</td>
</tr>
<tr>
<td>Gaussian</td>
</tr>
<tr>
<td>Uniform</td>
</tr>
</tbody>
</table>

B. Galerkin Projection for Linear DAE systems

For linear systems described by a set of ordinary differential equations (ODEs), Galerkin projection can be used to generate a set of deterministic ODEs to determine the PCE coefficients (e.g., [20], [21], [22]). This section outlines how to use Galerkin projection to generate a set of deterministic DAEs for determining the expansion coefficients, which provides an approach to efficiently compute state/output distributions for systems whose dynamics involve uncertain algebraic equations/constraints.

For the uncertain system (1), let \( x_i \) and \( r_i \) denote the \( i \)th component of \( x \) and \( r \), respectively, and \( M_{ij}, A_{ij}, \) and \( B_{ij} \) denote the \( ij \)th elements (i.e., \( i \)th row and \( j \)th column) of matrices \( M, A, \) and \( B \), respectively. Applying the truncated PCE expression (4) to each element in (1) gives

\[
\dot{x}_i (t, \theta) := \sum_{k=0}^{L} x_{ik}(t) \Phi_k(\theta) = x_i^T(t) \Lambda(\theta)
\]

\[
\dot{M}_{ij}(\theta) := \sum_{k=0}^{L} m_{ijk} \Phi_k(\theta) = m_{ij}^T \Lambda(\theta)
\]

\[
\dot{A}_{ij}(\theta) := \sum_{k=0}^{L} a_{ijk} \Phi_k(\theta) = a_{ij}^T \Lambda(\theta)
\]

\[
\dot{B}_{ij}(\theta) := \sum_{k=0}^{L} b_{ijk} \Phi_k(\theta) = b_{ij}^T \Lambda(\theta)
\]

\[
\dot{r}_i(\theta) := \sum_{k=0}^{L} r_{ik} \Phi_k(\theta) = r_i^T \Lambda(\theta)
\]

where \( x_i(t), m_{ij}, a_{ij}, b_{ij}, \) and \( r_i \in \mathbb{R}^{L+1} \) are defined similarly to \( a \) in (4). The elements of \( M, A, \) and \( B \) are known \( a \) priori and, therefore, their PCE coefficients can be computed from the normal equations

\[
m_{ijk} = \frac{\langle M_{ij}(\theta), \Phi_k(\theta) \rangle}{\langle \Phi_k(\theta)^2 \rangle}
\]

\[
a_{ijk} = \frac{\langle A_{ij}(\theta), \Phi_k(\theta) \rangle}{\langle \Phi_k(\theta)^2 \rangle}
\]

\[
b_{ijk} = \frac{\langle B_{ij}(\theta), \Phi_k(\theta) \rangle}{\langle \Phi_k(\theta)^2 \rangle}
\]
$$r_{ik} = \frac{\langle r_i(\theta), \Phi_k(\theta) \rangle}{\langle \Phi_k(\theta)^2 \rangle}.$$ 

There are a total of $n_x(L+1)$ unknown PCE coefficients for the states $\{x_{ik}\}_{k=0,\ldots,L}$. A deterministic DAE for these coefficients can be constructed by

1) Creating the polynomial chaos approximation of (1) by substituting the PCEs (7), (8), (9), (10), and (11) into (1) to give

$$\sum_{j=1}^{n_x} \sum_{l=0}^{L} m_{ijk} x_l \Phi_j \Phi_l$$

$$= \sum_{j=1}^{n_x} \sum_{l=0}^{L} a_{ijk} x_l \Phi_j \Phi_l + \sum_{j=1}^{n_x} \sum_{l=0}^{L} b_{ijk} u_j \Phi_k + \sum_{k=0}^{L} r_{ik} \Phi_k.$$ 

2) Projecting this approximation onto the orthogonal basis functions (i.e., $\{\Phi_k(\theta)\}$ for $i = 1, \ldots, n_x$ and $m = 0, \ldots, L$), which is optimal in the $L^2$ sense [19],

$$M X(t) = AX(t) + Bu(t) + R$$

where $X = [x_0^T \ x_1^T \ \cdots \ x_{n_x}^T]^T$, $R = [(Pr_0)^T (Pr_1)^T \ \cdots \ (Pr_{n_x})]^T$, and the matrices $M$, $A$, and $B$ are defined by their blocks,

$$M_{ij} = \sum_{k=0}^{L} m_{ijk} T_k$$

$$A_{ij} = \sum_{k=0}^{L} a_{ijk} T_k$$

$$B_{ij} = Pb_{ij}$$

with the symmetric inner product matrices $P$ and $T_k$ being

$$P = \begin{bmatrix} \langle \Phi_0^2 \rangle & 0 & \cdots & 0 \\
0 & \langle \Phi_1^2 \rangle & \cdots & 0 \\
\vdots & \vdots & \ddots & \vdots \\
0 & 0 & \cdots & \langle \Phi_L^2 \rangle \end{bmatrix},$$

$$T_k = \begin{bmatrix} \langle \Phi_0 \Phi_0 \Phi_0 \rangle & \langle \Phi_0 \Phi_0 \Phi_1 \rangle & \cdots & \langle \Phi_0 \Phi_0 \Phi_L \rangle \\
\langle \Phi_0 \Phi_1 \Phi_0 \rangle & \langle \Phi_0 \Phi_1 \Phi_1 \rangle & \cdots & \langle \Phi_0 \Phi_1 \Phi_L \rangle \\
\vdots & \vdots & \ddots & \vdots \\
\langle \Phi_0 \Phi_L \Phi_0 \rangle & \langle \Phi_0 \Phi_L \Phi_1 \rangle & \cdots & \langle \Phi_0 \Phi_L \Phi_L \rangle \\
\langle \Phi_1 \Phi_0 \Phi_0 \rangle & \langle \Phi_1 \Phi_0 \Phi_1 \rangle & \cdots & \langle \Phi_1 \Phi_0 \Phi_L \rangle \\
\vdots & \vdots & \ddots & \vdots \\
\langle \Phi_L \Phi_0 \Phi_0 \rangle & \langle \Phi_L \Phi_0 \Phi_1 \rangle & \cdots & \langle \Phi_L \Phi_0 \Phi_L \rangle \end{bmatrix}.$$ 

The PCE coefficients for the states (stacked into $X$) are composed of differential and algebraic states (see (1)). Denote the PCE coefficients of the differential and algebraic states as $X_d$ and $X_a$, respectively. Similar to (1), $X_a(0)$ and $X_d(0)$ are uniquely determined from $X_d(0)$ and (13) at $t=0$ (must be satisfied for consistent initialization). The elements of $X_a(0)$ can be computed from the known, but possibly uncertain, initial conditions $x_d(0, \theta)$ using

$$x_{ik}(0) = \frac{\langle x_i(0, \theta), \Phi_k(\theta) \rangle}{\langle \Phi_k(\theta)^2 \rangle}; \quad i = 1, \ldots, n_d$$

$$k = 0, \ldots, L$$

where the first $n_d$ elements of $x$ are $x_d$.

Note that (13) is a deterministic DAE in terms of the unknown time-varying PCE coefficients of the states. The PCE (7), along with the solution to (13), describes the uncertain states as an explicit function of time and $\theta$. This description can be used to efficiently compute the evolution of any order moment or cumulant of $\hat{z}(t, \theta)$ over time (see e.g., (5) and (6)) by exploiting the orthogonality property of the multivariate basis functions. Next, the main principles of QDMC are introduced.

IV. QUADRATIC DYNAMIC MATRIX CONTROL

This section provides a brief overview of the QDMC algorithm [23], [11], [24] for further details). For notational simplicity, the algorithm is first described for single-input single-output (SISO) systems. Due to the linear dynamics of (1), the SISO framework can be straightforwardly extended to multi-input multi-output (MIMO) systems, which is briefly discussed in Section IV-E.

A. Step Response

QDMC uses a finite step response (FSR) model to predict the dynamic response of the output to changes in the input from a steady state (assuming a piecewise constant input),

$$y_k = y_{ss} + \sum_{i=1}^{N-1} s_i \Delta u_{k-i} + s_N (u_{k-N} - u_{ss}),$$

(14)

where $y_k$ ($u_k$) denotes the model output (input) at discrete time instant $k$ (related to time $t$ via $t = T_s k$, where $T_s$ is the sampling time), $\Delta u_k = u_k - u_{k-1}$ denotes the change in the input at discrete time instant $k$, $y_{ss}$ ($u_{ss}$) denotes the initial steady-state output (input), $s_i$ denotes the $i$th step response coefficient caused by a unit step change in the input at time 0, and $N$ is the model length of the system (i.e., number of discrete time instants required for the system to reach the new steady state after the input step change at time 0). Assuming (1) represents a linear SISO system with fixed parameter values, the step response coefficients can be computed once off-line by performing a step test in the input and measuring the change in the output from its steady state. For (14) to be valid, the system must be asymptotically stable such that $s_N \approx s_{N+1} \approx \cdots \approx s_{\infty}$.

B. Prediction Model

The $p$-step ahead prediction of the output dynamic response can be expressed in terms of the FSR model (14),

$$\tilde{y}_{k+1} = T f_k + w_{k+1} + G \Delta u_k,$$

(15)

where

$$\tilde{y}_{k+1} = [\tilde{y}_{k+1}(k) \ \tilde{y}_{k+2}(k) \ \cdots \ \tilde{y}_{k+p}(k)]^T,$$

(16)

$$f_k = [f_{k} f_{k+1} \cdots f_{k+N-1}]^T,$$

(17)

$$w_{k+1} = [w_{k+1}(k) \ w_{k+2}(k) \ \cdots \ w_{k+p}(k)]^T,$$

(18)

$$\Delta u_k = [\Delta u_{k} \ \Delta u_{k+1} \ \cdots \ \Delta u_{k+m-1}]^T,$$

(19)

$x_{ik}$ denotes the value of $x$ at time instant $i$ given information up to and including time instant $k$, $\tilde{y}_{k+1}$ denotes the predicted outputs over the prediction horizon $p$, $f_k$ is the free
response of the system (i.e., output response over the model length \( N \) assuming that current and future input changes are zero), \( w_{k+1} \) denotes the future unmeasured disturbances, and \( \Delta u_k \) denotes the future inputs over the control horizon \( m < p \). The binary matrix \( T \in \mathbb{R}^{p \times N} \) displaces \( f_k \), leaving only the elements \( f_{k+1:k} \) to \( f_{k+p:k} \) needed for prediction. The dynamic matrix \( G \in \mathbb{R}^{p \times m} \) describes how current and future input changes will affect the system output, and is composed of the step response coefficients,

\[
G = \begin{bmatrix}
  s_1 & 0 & 0 & \cdots & 0 & 0 \\
  s_2 & s_1 & 0 & \cdots & 0 & 0 \\
  \vdots & \vdots & \ddots & \ddots & \vdots & \vdots \\
  s_j & s_{j-1} & s_{j-2} & \cdots & s_{j-m+1} & 0 \\
  \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\
  s_p & s_{p-1} & s_{p-2} & \cdots & s_{p-m+1} & 0 \\
\end{bmatrix}.
\]

The free response vector is updated with the new input applied at time \( k \) by

\[
f_{k+1} = \begin{bmatrix}
  0 & 1 & 0 & \cdots & 0 & 0 \\
  0 & 1 & \cdots & \cdots & \cdots & \cdots \\
  \vdots & \vdots & \ddots & \ddots & \vdots & \vdots \\
  0 & \cdots & \cdots & \cdots & 0 & 1 \\
  0 & \cdots & \cdots & \cdots & \cdots & 1 \\
\end{bmatrix} f_k + \begin{bmatrix}
  s_1 \\
  s_2 \\
  s_3 \\
  \vdots \\
  s_{N} \\
\end{bmatrix} \Delta u_k,
\]

which is essential for the receding horizon implementation of QDMC.

C. Disturbance Update

The model (15) cannot be directly used for prediction as the future, unmeasured disturbances \( w_{k+1} \) are unknown. To incorporate the effect of unmeasured disturbances, QDMC makes the additive disturbance assumption to compute an estimate of \( w_{k+1} \), which is a valid assumption for systems that integrate all unmeasured disturbances [25]. The additive disturbance assumption indicates that

1) The unmeasured disturbance at time \( k \) (\( w_{k|k} \)) can be estimated as the difference between the current measurement \( y_k \) and the predicted output \( \hat{y}_{k|k} = f_{k|k} \), i.e., \( w_{k|k} = y_k - \hat{f}_{k|k} \).

2) The unmeasured disturbance remains constant into the future, i.e., \( w_{k|k} = w_{k+1|k} = \cdots = w_{k+p|k} \).

Using the additive disturbance assumption, (15) can be rewritten to obtain the prediction model used in QDMC,

\[
\begin{align*}
\bar{y}_{k+1} & = \begin{bmatrix}
Tf_k + \mathcal{I} (y_k - \hat{f}_{k|k}) + G \Delta u_k
\end{bmatrix} \\
\end{align*}
\]

(20)

where \( \mathcal{I} = \begin{bmatrix} 1 & 1 & \cdots & 1 \end{bmatrix}^T \in \mathbb{R}^p \). Henceforth, the past (free response) and present (feedback) terms will be grouped into a single vector denoted \( \bar{y}_{k+1} = Tf_k + \mathcal{I} (y_k - \hat{f}_{k|k}) \) for notational convenience.

D. Control Formulation via Optimization

The control objective in QDMC is to determine an optimal input sequence that results in the system output being as close as possible to a desired reference trajectory \( r \), while satisfying input and output constraints [24], [26]:

\[
\begin{align*}
\min_{\Delta u_k} & \ (y_{k+1} - r)^T W_y (y_{k+1} - r) + \Delta u_k^T W_u \Delta u_k \\
\text{s.t.} & \ u_{\text{min}} \leq u_k \leq u_{\text{max}} \\
& \Delta u_{\text{min}} \leq \Delta u_k \leq \Delta u_{\text{max}} \\
& y_{\text{min}} \leq y_{k+1} \leq y_{\text{max}}
\end{align*}
\]

where \( W_y \) is the output weight, \( W_u \) is the input change weight, \( u_{\text{min}} \) (\( u_{\text{max}} \)) is the minimum (maximum) allowable input, \( \Delta u_{\text{min}} \) (\( \Delta u_{\text{max}} \)) is the minimum (maximum) allowable change in input, and \( y_{\text{min}} \) (\( y_{\text{max}} \)) is a vector of the smallest (largest) possible output values at each time over the prediction horizon. Eq. (21) can be rewritten as a quadratic program (QP) in \( \Delta u_k \) by substituting (20) (e.g., see [24]).

E. Extension to MIMO Systems

For a linear system with \( n_y \) outputs and \( n_u \) inputs, a MIMO prediction model can be obtained by combining all the SISO prediction models for different input-output pairs

\[
\begin{bmatrix}
\bar{y}_{k+1}^1 \\
\vdots \\
\bar{y}_{k+1}^{n_y}
\end{bmatrix} = \begin{bmatrix}
G^{1,1} & \cdots & G^{1,n_u} \\
\vdots & \ddots & \vdots \\
G^{n_y,1} & \cdots & G^{n_y,n_u}
\end{bmatrix} \begin{bmatrix}
\Delta u_{k}^1 \\
\vdots \\
\Delta u_{k}^{n_u}
\end{bmatrix}
\]

where \( \bar{y}_{k+1}^i \) denotes the \( i \)th predicted output, \( \bar{y}_{k+1}^{p_i,i} \) denotes the feedback corrected free response corresponding to the \( i \)th output, \( \Delta u_{k}^i \) denotes the \( i \)th input, and \( G^{i,j} \) is the dynamic matrix of the SISO system composed of the \( i \)th output and \( j \)th input. This equation has the same form as the SISO prediction in (20), which implies that the formulation of the control problem (21) also holds for MIMO systems. In this case, the control weights \( W_y \) and \( W_u \) should be defined as positive semi-definite matrices to enable the various outputs/inputs of the MIMO system have different weightings in the control objective.

V. Fast Stochastic MPC

This section presents a computationally tractable SMPC formulation for Problem 1 that is particularly suited to high-dimensional systems. The input-output framework of QDMC (see Sec. IV) is exploited to render the stochastic control approach independent of the state dimension, while PCEs with the Galerkin projection method adapted for DAE systems (see Sec. III) is used for efficient propagation of probabilistic uncertainties. In addition, the SMPC approach eliminates the need for state estimation owing to the disturbance update step of QDMC, which is based on the output measurements.

In this fast SMPC approach, the control objective function \( J \) is defined in terms of the expected value and variance of the output PDFs. Eq. (13) is used to propagate the uncertainties \( \theta \) through the system dynamics, which generates a set of linear DAEs in terms of the unknown state PCE.

2806
coefficients $X(t)$. For the PCE coefficients of the outputs denoted by $Y(t)$, linearity of (13) implies that FSR models can be developed for the elements of $Y(t)$, which provide the output PCE coefficients as a function of the system inputs. Then, moments of the output PDFs can be derived in terms of $Y(t)$ (e.g., (5) and (6)) to state the SMPC optimization problem as a QP in the inputs. For notational simplicity, the fast SMPC algorithm is presented for SISO systems, which can be straightforwardly extended to MIMO systems as described in Sec. IV-E.

A. Objective

A multi-objective optimization in terms of the moments of the output distribution avoids the conservatism of robust (worst-case) control [27] while shaping the distribution to have a desired PDF [9]. Due to space constraints, only the first two central moments are detailed here. However, it is straightforward to extend the results for higher moments, which can still be efficiently approximated using PCE (e.g., [19], [20]). The proposed objective, over the finite horizon $p$, is

$$J = (E[y_{k+1}] - r_e)^T W_e (E[y_{k+1}] - r_e) \quad (22)$$

$$+ W_v \sum_{i=1}^p \text{Var}[\tilde{y}_{k+i|k}] + \Delta u_k^T W_u \Delta u_k$$

where $y_{k+1}$ is defined in (16), $\Delta u_k$ is defined in (19), $r_e$ is the setpoint trajectory for the expected value, $W_e$ is the output expected value weight, and $W_v$ is the output variance weight. In (22), the variance is summed over the prediction horizon, which is a plausible objective function since the variance must be greater than or equal to zero by definition, and allows the optimization to be written as a QP as detailed in the subsequent sections. If the nominal output trajectory were used in (22) instead of the expected value, then the subsequent results also hold with minor modifications.

B. Prediction Model for Output PCE

Propagating the uncertain variables $\theta$ through the linear DAE dynamics in (1) produces a deterministic DAE (13) for the unknown state PCE coefficients $X(t)$, which is linear in the input $u(t)$. This linearity implies that an FSR model can be developed for the every PCE coefficient of the state (i.e., every element of $X(t)$) by simply performing a unit step test on the input (Section IV-A). Assuming a SISO system, let the output-truncated PCE be defined as

$$\hat{y}(t, \theta) := \sum_{j=0}^L y_j(t) \Phi_k(\theta) \quad (23)$$

where $y_j(t), \forall j = 0, \ldots, L$, are the output PCE coefficients. Since the output is lumped into the state, performing the unit step in the input directly provides FSR models for $y_j$. The $p$-step ahead prediction for $y_j$ can be constructed similarly to (15) (without the unmeasured disturbance $w_{k+1}$):

$$\tilde{y}_{j,k+1} = T_f y_j + G_j \Delta u_k \quad \forall j = 0, \ldots, L, \quad (24)$$

where $\tilde{y}_{j,k+1}$ refers to the predicted $j$th PCE coefficient of the output evaluated discretely over the horizon from $k + 1$ to $k + p$.

C. QP Formulation

This section reformulates Problem 1 as a QP for $\Delta u_k$ by substituting the PCE coefficient prediction model (24) into the objective (22) and the input/output constraints. The expected value is simply the first coefficient of the output PCE coefficients as a function of the system inputs.

The open-loop control problem is obtained by neglecting the measurement update in this expected value term. Open-loop formulations can be useful when measurements are not available or reliable.

Using (6), the variance over the prediction horizon can be rearranged as

$$\sum_{i=1}^p \text{Var}[\tilde{y}_{k+i|k}] = \sum_{i=1}^p \sum_{j=1}^L \tilde{y}_{j,k+i|k}^2 \langle \Phi_j^2 \rangle = \sum_{j=1}^L \langle \Phi_j^2 \rangle \tilde{r}_{j,k+1}$$

where

$$
\tilde{r}_{j,k+1} = \sum_{i=1}^L \Phi_j^T W_e \Phi_j \tilde{y}_{j,k+1}$$

Substituting (24), (25), and (26) into (22) converts $J$ into an explicit quadratic function of $\Delta u_k$. A similar procedure can be used to write the input, actuator (change-in-input), and output inequality constraints as explicit linear functions of $\Delta u_k$. The details of these manipulations are omitted due to space constraints. The final QP reformulation of Problem 1 is stated below.

Problem 2. (Fast SMPC for high-dimensional systems):

$$\begin{align*}
\min_{\Delta u_k} & \quad J = \frac{1}{2} \Delta u_k^T H \Delta u_k + e^T \Delta u_k \\
\text{s.t.} & \quad A_c \Delta u_k \leq b_c
\end{align*} \quad (27)$$

where

$$H = G_0^T W_e G_0 + \sum_{j=1}^L \langle \Phi_j^2 \rangle \sum_{i=1}^p \tilde{y}_{j,k+i|k}^2 \Phi_j^T W_e \Phi_j$$

$$c = G_0^T W_e (y_{0,k+1} - r_e) + \sum_{j=1}^L \langle \Phi_j^2 \rangle \sum_{i=1}^p \tilde{r}_{j,k+1}^2 \Phi_j^T W_e \Phi_j y_{j,k+1}^p$$

$$y_{j,k+1}^p = \begin{cases} T_f y_j + I(y_k - f_{0,k}) & \text{if } j = 0, \\
T_f y_j & \text{if } j \in \{1, \ldots, L\}, 
\end{cases} \quad (28)$$
\[ A_c = \begin{bmatrix} I & -I & 0 & \cdots & 0 \\ -I & I & -I & \cdots & 0 \\ \vdots & \vdots & \ddots & \ddots & \vdots \\ 0 & 0 & \cdots & I & -I \\ 0 & 0 & \cdots & -I & I \end{bmatrix}, \quad b_c = \begin{bmatrix} (u_{\text{max}} - u_{k-1})I_1 \\ -(u_{\text{min}} - u_{k-1})I_1 \\ \Delta u_{\text{max}}I_1 \\ -\Delta u_{\text{min}}I_1 \\ y_{\text{max}}(\theta) - \hat{y}^p_{k+1}(\theta) \\ -y_{\text{min}}(\theta) + \hat{y}^p_{k+1}(\theta) \end{bmatrix} \]

The concepts of quadratic dynamic matrix control (QDMC) are introduced for systems with time-invariant probabilistic uncertainties. The linearity of the proposed approach allows for the use of linear control techniques, which are shown to be effective in high-dimensional systems with process uncertainties and disturbances.

In this work, the nonlinear DAE model is linearized around a desired steady-state operating condition and, subsequently, the linearized model is used to develop a fast SMPC controller (from Problem 2) to suppress the adverse effects of uncertainties in kinetic parameters on the critical quality attributes (CQAs) of tablets. The objective is stated as a setpoint tracking problem for the production rate and API with an upper bound on the impurity content of the tablets included as a constraint. The performance of the fast SMPC controller is evaluated for the case of a 5\% step increase in the production rate. Fig. 1 shows the production rate for the fast SMPC and QDMC for 200 closed-loop simulations, where the step change in the setpoint is applied at 5 hr. The fast SMPC controller results in a lower variance in the production rate than that of the QDMC controller (Fig. 1), providing a more robust performance in the presence of system uncertainties.

The SMPC approach leads to tighter setpoint tracking of the API dosage profiles as well (see Fig. 2). Fig. 3 shows the distributions of the API dosage at different points in the course of the transient system dynamics resulting from the production rate setpoint change. The SMPC produces a lower variance in the API dosage at all times. The variance in the API dosage at time 65 hr is a factor of 25 lower for the fast SMPC controller than that of the QDMC controller (Table II). The optimization (27) consistently took less than 1 second to solve (on a laptop running Windows 7 with 8 GB of RAM) for a variety of prediction and control horizons, meaning that the control inputs could easily be found and supplied to the plant in real-time. Given the high importance of drug products meeting the specifications in the presence of uncertainties, these results indicate that SMPC is a promising approach for application to pharmaceutical manufacturing. Moreover, the speed at which the proposed fast MPC algorithm is able to compute the solution indicates that algorithm could easily be implemented in real-time on large and complex processes.

VI. STOCHASTIC CONTROL OF A CONTINUOUS PHARMACEUTICAL MANUFACTURING PROCESS

This paper presents a stochastic MPC approach for high-dimensional systems with time-invariant probabilistic uncertainties. The concepts of quadratic dynamic matrix control
and polynomial chaos expansion are used to develop an input-output formulation for fast SMPC whose on-line computational cost is independent of the state dimension. This approach circumvents the prohibitive online computational costs associated with model predictive control of uncertain high-dimensional systems. The Galerkin projection is modified to handle a general class of linear DAEs so that the SMPC approach is applicable to a large class of complex systems that include descriptor/singular systems. The effectiveness of the proposed control approach is demonstrated for control of a continuous pharmaceutical manufacturing process with nearly 8000 states, where the SMPC controller effectively regulates the critical quality attributes of the product in the presence of parametric uncertainties. In addition, the proposed control law (computed using the optimization in (27)) was found in less than 1 second at each time step so that the algorithm can easily be implemented in real-time. This work is being extended to incorporate joint chance constraints into the fast SMPC approach.

**REFERENCES**


