Preprints of the 19th World Congress
The International Federation of Automatic Control
Cape Town, South Africa. August 24-29, 2014

Optimal Experimental Design for Probabilistic Model Discrimination Using Polynomial Chaos

Stefan Streif * **Felix Petzke** * **Ali Mesbah** ** **Rolf Findeisen** * **Richard D. Braatz** **

* *Otto-von-Guericke-University Magdeburg, Germany; e-mail:* {stefan.streif, rolf.findeisen}@ovgu.de, {felix.petzke}@me.com.
** *Massachusetts Institute of Technology, Cambridge, MA, USA; e-mail:* {amesbah,braatz}@mit.edu.

Abstract: Building dynamic models is important in many applications including model-based design, optimization, and control. When multiple hypothesized models have predictions that are consistent with the measurements, experimental design is used to discriminate between the models. This task is particularly challenging for nonlinear systems subject to uncertainties. An optimal experimental design method for model discrimination for polynomial uncertain systems is presented that can be used to discriminate models based on dissimilarity of the probability densities of the model outputs. Generalized polynomial chaos theory in conjunction with Galerkin projection is used to derive an extended set of ordinary differential equations. Simulation of the extended system enables prediction of the propagation of probabilistic uncertainties associated with the model parameters and initial conditions, and to obtain the output probability densities. The simulation of the hypothetical models is embedded in a nonlinear optimization problem to determine an optimal input sequence that maximizes model dissimilarity. The experimental design method is demonstrated using a numerical example.

Keywords: Experimental design; nonlinear systems; probabilistic uncertainties; model discrimination.

1. INTRODUCTION

Dynamical models are widely utilized in engineering applications for system design and control purposes. However, often several hypothesized models are available for a given system and performing a large number of experiments for (in)validation of the different model hypotheses can be expensive. In addition, system nonlinearities as well as uncertainties associated with initial conditions, model parameters, and measurements make the model (in)validation task challenging. For example, in chemical and biological systems, the uncertainties that result from variability of physiochemical phenomena or experimental variations often lead to probabilistically distributed model outputs. Therefore, methods are desired to systematically design experiments that, with high probability, discriminate between competing nonlinear models subject to probabilistic uncertainties.

Experimental design for model discrimination has been extensively investigated (e.g., see Fedorov (1972), Goodwin and Payne (1977), Zarrop (1979), Walter and Pronzato (1990), Pukelsheim and Rosenberger (1993), Chen and Asprey (2003), Borchers et al. (2011), and the references therein), where auxiliary input signals are designed to facilitate the separation of multiple models. A traditional approach to determine the discrimination criterion is based on Bayesian inference, in which prior probabilities are associated with the competing models (Box and Hill, 1967). Input signal design for model discrimination is then performed using the predicted posterior probability computed from Bayes' theorem. A more widely used approach is based on the so-called *alphabetic-optimality criteria*, where, assuming that one model is correct, the difference between the predictions of the competing models is maximized (Atkinson and Fedorov, 1975). Related work on active fault diagnosis (e.g., Campbell and Nikoukhah (2004), Blanke et al. (2006), Scott et al. (2013), Mesbah et al. (2014a), Paulson et al. (2014)) deal with input signal design for fault model invalidation.

To address the experimental design problem for uncertain systems, a relaxation approach has been presented by Georgiev and Klavins (2008) to derive disparity certificates for stochastic inputs that yield different outputs for all possible disturbances; hence, enabling model discrimination in the presence of uncertainties. Halder and Bhattacharya (2012) proposed a probabilistic formulation for model validation using the Wasserstein metric and Liouville's equation for uncertainty propagation.

This paper considers polynomial ordinary differential equations, which often arise in the modeling of (bio)chemical reaction networks or can be obtained by approximation or state immersion (Ohtsuka, 2005). An op-

Corresponding author: Stefan Streif.

timal experimental design method is presented to discriminate uncertain nonlinear models based on dissimilarity of the probability density of the model outputs.

The problem of optimal experimental design is formally stated in Sec. 2. In Sec. 3, the probabilistic uncertainties associated with model parameters and initial conditions are propagated using generalized Polynomial Chaos (PC) theory (Wiener, 1938; Xiu and Karniadakis, 2002) in conjunction with Galerkin projection. The resulting higher-dimensional set of deterministic ordinary differential equations (ODEs) is used to construct the probability density functions (PDFs) of the outputs. Different strategies to construct the PDF of a stochastic variable needed for probabilistic model discrimination are discussed and a computationally efficient approach for the approximation of unimodal PDF is proposed. Sec. 4 presents a measure for dissimilarity of the PDFs of the model outputs. The optimal input design for experimental design is posed in form of a nonlinear optimization problem minimizing an input norm, while at the same time enforcing an upper bound on the overlap of the noise-corrupted output PDFs as determined by PC. Sec. 5 contains a numerical illustration of the proposed method.

Notation. n_z (z_i) denotes the number of elements (resp. the i^{th} element) of a vector z. $\Phi_{(i)}^{(m)}$ represents a univariate polynomial of order m of stochastic variables ξ_i, $i \in \{1, \ldots, n_\xi\}$. $\{\Phi_{(i)}^{(m)}\}_{m=0}^{P}$ denotes a finite sequence of $P+1$ polynomials of order up to P, which are orthonormal with respect to the probability measure $\mu(\xi_i)$. The Kronecker product is represented by $\otimes$, and $\bigotimes_{i=1}^{n} v_i$ denotes the sequence of Kronecker products $v_1 \otimes v_2 \otimes \cdots \otimes v_n$. $\nu^{(m)}(v)$ denotes the m^{th} moment of a random variable or stochastic variable v. $\langle \phi_1(\xi), \phi_2(\xi) \rangle$ denotes the inner product $\int_{\operatorname{spt}\xi} \phi_1(\xi)\phi_2(\xi)\mu(d\xi)$ with respect to the probability measure $\mu(\xi)$ with support $\operatorname{spt}\xi$. The number of times that element i appears in list A is denoted by $\#(A, i)$.

2. PROBLEM FORMULATION

This paper considers nonlinear dynamical models of the form

$$m^{[j]} : \begin{cases} \dot{x}^{[j]}(t) = f^{[j]}\big(x^{[j]}(t), p^{[j]}, u^{[j]}(t)\big) \\ y^{[j]}(t) = h^{[j]}\big(x^{[j]}(t), p^{[j]}, u^{[j]}(t)\big) \end{cases} \quad j \in \{1, 2\} \tag{1}$$

where the superscript $[j]$ denotes the model index and the variables of the corresponding model; the variables $x^{[j]} \in \mathbb{R}^{n_x}$, $u^{[j]} \in \mathbb{R}^{n_u}$, $y^{[j]} \in \mathbb{R}^{n_y}$, and $p^{[j]} \in \mathbb{R}^{n_p}$ denote the states, inputs, outputs, and parameters, respectively; and $f^{[j]}$ and $h^{[j]}$ denote polynomial functions. To shorten the notation, the dimensions of the variable vectors are assumed to be the same for all models. This paper formulates the optimal experimental design problem for only two models, both due to space limitations and because the extension to multiple models is straightforward.

Probabilistic uncertainties associated with model parameters and initial conditions are considered. Denote the vector of independent random variables by $\xi^{[j]} := [p_1^{[j]}, \ldots, p_{n_p}^{[j]}, x_1^{[j]}(0), \ldots, x_{n_x}^{[j]}(0)]^\mathsf{T} \in \mathbb{R}^{n_\xi}, n_\xi = n_p + n_x$.

To discriminate the two models, it is assumed that experiments are performed over a finite time horizon $[0, T]$ and m_y measurements are taken at discrete time points $t_{y,k} \in [0, T], k = 1, \ldots, m_y$. The measurements $\widehat{y}_{lk}^{[j]} := \widehat{y}_l^{[j]}(t_{y,k})$, $l = 1, \ldots, n_y$ are expressed by noise-corrupted model outputs as

$$\widehat{y}_{lk}^{[j]} = y_{lk}^{[j]}\left(1 + w_{lk}^{[j]}\right), \quad l = 1, \ldots, n_y, \quad j \in \{1, 2\}, \tag{2}$$

where $y_{lk}^{[j]}$ and $w_{lk}^{[j]}$ are stochastically independent. The measurement noise $w_i^{[j]}(t_{y,k}) \in \mathbb{R}$ is assumed to be multiplicative with known PDF $\mu(w_{lk}^{[j]})$.

The probabilistic uncertainties of the parameters, initial conditions, and measurement noise lead to probabilistically distributed outputs with the PDFs $\mu(\widehat{y}_i^{[j]}(t))$. Model discrimination is largely hindered by the distributed character of the outputs. If the support of the PDFs $\mu(\widehat{y}_i^{[1]}(t))$ and $\mu(\widehat{y}_i^{[2]}(t))$ overlap and have non-negligible probabilities on the support intersection, measurements may not allow unambiguous invalidation of one of the models. The model invalidation becomes even more difficult due to the measurement noise.

This paper aims to design an input sequence to be applied to the true system such that actual measurements can be associated, with a high probability, either with the noise-corrupted outputs of model $m^{[1]}$ or of model $m^{[2]}$.A control vector parameterization with piecewise constant input sequence, which is changed at m_u different time-points $t_{u,k} \in [0, T], k = 1, \ldots, m_u$, is adopted,

$$u(t) = \begin{cases} \bar{u}(t_{u,1}), & t < t_{u,2} \\ \bar{u}(t_{u,k}), & t_{u,k} \le t < t_{u,k+1}, k = 2, \ldots, m_u - 1 \\ \bar{u}(t_{u,m_u}), & t \ge t_{u,m_u} \end{cases}$$

where the input values are restricted to $\mathcal{U} \subseteq \mathbb{R}^{n_u}$. The input design problem is stated as follows.

Problem 1 (Experimental Design for Model Discrimination): *Find an optimal input sequence $\{\bar{u}(t_{u,1}), \ldots, \bar{u}(t_{u,m_u})\}$ to both models $m^{[1]}$ and $m^{[2]}$ (i.e., $u(t) = u^{[1]}(t) = u^{[2]}(t)$) such that the output PDFs are separated at least at one (measurement) time instant in the presence of measurement noise.*

To solve Problem 1, three subproblems should be addressed. First, the probability density functions of the model outputs resulting from the probabilistic uncertainties of model parameters and initial conditions should be constructed, which requires propagation of the probabilistic uncertainties through the nonlinear model (Sec. 3). Second, a measure of the dissimilarity of the output PDFs is required. Third, a nonlinear optimization problem should be formulated to determine an optimal input sequence that minimizes the overlap between the output PDFs. The latter two subproblems are discussed in Sec. 4.

3. UNCERTAINTY PROPAGATION USING POLYNOMIAL CHAOS

This section deals with the propagation of the probabilistic uncertain parameters and initial conditions through the nonlinear dynamical models. For this purpose, a Galerkin-based Polynomial Chaos (PC) approach is employed (e.g., see Kim et al. (2013)). In the Galerkin-based PC approach, the original nonlinear system (1) is transformed into a nonlinear system of ordinary differential equations of higher dimension, which is then solved to determine the coefficients of the PC expansions. The moments of the stochastic variables can be computed based on the PC coefficients. Besides the Galerkin method, collocation approaches can also be used to determine the coefficients of the PC expansions (e.g., see Mesbah et al. (2014b) and the citations therein). To simplify the notation in this section, the superscripts $[j], j \in \{1,2\}$ on the variables are dropped.

Consider the polynomial dynamics for the r^{th} state,

$$\dot{x}_r = f_r(x,p,u) = \sum_{i=1}^{o_r} p_{ji} \prod_{l=1}^{n_x} x_l^{\alpha_{ril}} \prod_{q=1}^{n_u} u_q^{\beta_{riq}}, \qquad \forall r = 1,\ldots,n_x, \quad (3)$$

where o_r denotes the number of monomials and the i^{th} monomial has total degree $\sum_{l=1}^{n_x} \alpha_{ril} + \sum_{q=1}^{n_u} \beta_{riq}$, where α_{ril} and β_{riq} denote the degree of variables x_l and u_q, respectively. The polynomial output map h in (1) can be represented similarly.

The PC approach is not limited to polynomial systems. Polynomial dynamics, however, simplify the evaluation of the multidimensional integrals arising due to the Galerkin projection. In general it is not a limitation, as many nonlinear differentiable functions (e.g., rational, exponential, and transcendental) can be approximated or exactly represented by polynomial dynamics using state immersion (Ohtsuka, 2005).

3.1 Polynomial Chaos Expansion

All stochastic variables v affected by the probabilistic uncertainties of the (stochastically independent) variables $\xi_i, i = 1,\ldots,n_\xi$ (i.e., the initial conditions and the model parameters, but not the input which is considered certain) can be approximated using a PC expansion of order P (Wiener, 1938; Xiu and Karniadakis, 2002)

$$v(\xi) \approx v_{(0)} + \sum_{i_1=1}^{n_\xi} v_{(i_1)} \Psi^{(1)}_{(i_1)} + \sum_{i_1=1}^{n_\xi} \sum_{i_2=1}^{i_1} v_{(i_1,i_2)} \Psi^{(2)}_{(i_1,i_2)} + \cdots + \sum_{i_1=1}^{n_\xi} \sum_{i_2=1}^{i_1} \cdots \sum_{i_P=1}^{i_{P-1}} v_{(i_1,\ldots,i_P)} \Psi^{(P)}_{(i_1,\ldots,i_P)}, \quad (4)$$

where $v_{(i_1,i_2,\ldots)}$ denote the coefficients of the PC expansion that describe the influence of the uncertainties of the variables $\xi_{i_1}, \xi_{i_2}, \ldots$ on the stochastic variable v. $\Psi^{(m)}_{(i_1,i_2,\ldots)}$ denotes the multivariate polynomial in the random variables $\xi_{i_1}, \xi_{i_2}, \ldots$ of total degree m. The multivariate polynomials can be written in terms of univariate polynomials: $\Psi^{(m)}_{(i_1,i_2,\ldots)} := \Phi^{(m_1)}_{(1)} \Phi^{(m_2)}_{(2)} \cdots \Phi^{(m_{n_\xi})}_{(n_\xi)}$, with $m = m_1 + \cdots + m_{n_\xi}$, and $m_i = \#(\{i_1, i_2, \ldots\}, i), i = 1,\ldots,n_\xi$. The polynomial bases $\{\Phi^{(\tilde{m})}_{(i)}\}_{\tilde{m}=0}^{P}$, $i = 1,\ldots,n_\xi$ are orthogonal with respect to the corresponding PDFs $\mu(\xi_i)$ of the random variable ξ_i. For standard distributions such as Normal or Beta distributions, orthogonal bases are readily available (Xiu and Karniadakis, 2002) or, in general, the orthogonal bases can be constructed based on moments (Oladyshkin and Nowak, 2012) or using Gram-Schmidt orthogonalization (Gerritsma et al., 2010).

Eq. (4) corresponds to a weighted sum of polynomials with $\widetilde{P} := \frac{(n_\xi+P)!}{n_\xi! P!}$ terms in total and can be written in a compact form as

$$v(\xi) \approx \widetilde{v}^{\mathsf{T}} \widetilde{\Phi},$$

with

$$\widetilde{v} := [v_{(0)}, v_{(1)}, v_{(2)}, \ldots, v_{(n_\xi)}, v_{(1,1)}, v_{(1,2)}, \ldots, v_{(1,1,1)}, v_{(1,1,2)}, \ldots, v_{(n_\xi,n_\xi,\cdots,n_\xi)}]^{\mathsf{T}} \in \mathbb{R}^{\widetilde{P}} \quad (5)$$

being the vector of coefficients of the PC expansion and

$$\widetilde{\Phi} := [1, \Phi^{(1)}_{(1)}, \Phi^{(1)}_{(2)}, \ldots, \Phi^{(1)}_{(n_\xi)}, \Phi^{(2)}_{(1)}, \Phi^{(1)}_{(1)}\Phi^{(1)}_{(2)}, \ldots, \Phi^{(3)}_{(1)}, \Phi^{(2)}_{(1)}\Phi^{(1)}_{(2)}, \ldots, \Phi^{(P)}_{(n_\xi)}]^{\mathsf{T}} \in \mathbb{R}^{\widetilde{P}}$$

being a $\widetilde{P}$-dimensional vector of the multivariate polynomials.

3.2 Galerkin Projection

To obtain the coefficients of the PC expansions, an extended system of ordinary differential equations is derived using Galerkin projection (Wiener, 1938; Xiu and Karniadakis, 2002). The system of ODEs describes the dynamics of the coefficients. Inserting the PC expansion of the stochastic variables (4) into the system dynamics (3) results in

$$\dot{\widetilde{x}}_r = \widetilde{f}_r(\widetilde{x}, \widetilde{p}, u, \widetilde{\Phi}) = \sum_{i=1}^{o_r} E_{\tilde{\alpha}_{ri}} \left(\widetilde{p}_{ri} \otimes \left(\bigotimes_{l=1}^{n_x} \bigotimes_{a=1}^{\alpha_{ril}} \widetilde{x}_l \right) \prod_{q=1}^{n_u} u_q^{\beta_{riq}} \right) \qquad \forall r = 1,\ldots,n_x, \quad (6)$$

where $\widetilde{x}_r$ and $\widetilde{p}_{ri}$ are the vectors of coefficients of the PC expansions (cf. (5)) of the states and parameters, respectively. The system (6) has extended state space dimension $\widetilde{P} n_x$ and describes the dynamics of the PC expansion coefficients. The outputs $\widetilde{y}_r = \widetilde{h}_r(\widetilde{x}, \widetilde{p}, u, \widetilde{\Phi})$, $r = 1,\ldots,n_y$ can be represented similarly.

The projection matrix $E_{\tilde{\alpha}_{ri}}$ in (6) accounts for the Galerkin projections of the products of the different multivariate orthogonal polynomials due to the polynomial terms appearing in (6). The k^{th} row, $k = 1,\ldots,\widetilde{P}$ of $E_{\tilde{\alpha}_{ji}}$ is given by

$$[e_{k,11\cdots1}, e_{k,21\cdots1}, \ldots, e_{k,\widetilde{P}1\cdots1}, e_{k,12\cdots1}, \ldots, e_{k,1\widetilde{P}\cdots1}, \ldots, e_{k,\widetilde{P}\widetilde{P}\cdots\widetilde{P}}]^{\mathsf{T}} \in \mathbb{R}^{\widetilde{P}^{\tilde{\alpha}_{ri}+1}},$$

with $\tilde{\alpha}_{ri} := \sum_{l=1}^{n_x} \alpha_{ril}$ and

$$e_{k,i_1 i_2 \cdots i_p} = \frac{\langle \widetilde{\Phi}_k, \widetilde{\Phi}_{i_1} \cdots \widetilde{\Phi}_{i_p} \rangle}{\langle \widetilde{\Phi}_k, \widetilde{\Phi}_k \rangle}, \qquad k, i_1, \ldots, i_p \in \{1, \ldots, n_{\widetilde{P}}\}. \quad (7)$$

The majority of $e_{i_j,i_1 i_2 \cdots i_p}$ are zero, independent of the chosen basis, due to the properties of (power) orthogonal polynomials (Gautschi et al., 2004; Milovanović, 2001). These properties, as well as symmetries such as $e_{i_j,12} = e_{i_j,21}$, can be exploited to reduce the large computational burden of building (6). The projection integrals can be evaluated using Gauss Quadrature, which is exact for the considered polynomial dynamics.

Once the projection coefficients (7) have been determined, the initial conditions $\widetilde{x}(0)$ as well as the coefficients of the PC expansions corresponding to the parameters $\widetilde{p}_j$ have to be determined, which can be obtained by projection of the corresponding PC expansion (4) onto the different orthogonal polynomials $\widetilde{\Phi}_k$. For $v = \xi_i$, this projection is

$$\widetilde{v}_k = \frac{\langle \xi_i, \widetilde{\Phi}_k \rangle}{\langle \widetilde{\Phi}_k, \widetilde{\Phi}_k \rangle}, \quad k = 1, \dots, \widetilde{P}, \tag{8}$$

which is zero in most cases due to orthogonality.

3.3 Computation of the Moments

The system (6) and the initial conditions obtained from (8) can be simulated to obtain the moments of the PDFs of the model outputs, as required for formulating the optimal experimental design problem. The moments of a stochastic variable v can be derived based on the coefficients $\widetilde{v}$ of its respective PC expansion (Fisher and Bhattacharya, 2009)

$$\nu^{(1)}(v) := \int_{\operatorname{spt} \xi} v(\xi)\mu(d\xi) = \sum_{i_1=1}^{\widetilde{P}} \left(\widetilde{v}_{i_1} \langle 1, \widetilde{\Phi}_{i_1} \rangle \right) = v_{(0)}$$

$$\nu^{(2)}(v) := \int_{\operatorname{spt} \xi} v(\xi)^2 \mu(d\xi) = \sum_{i_1=1}^{\widetilde{P}} \sum_{i_2=1}^{\widetilde{P}} \left(\widetilde{v}_{i_1} \widetilde{v}_{i_2} \langle \widetilde{\Phi}_{i_1}, \widetilde{\Phi}_{i_2} \rangle \right)$$
$$= \sum_{i=1}^{\widetilde{P}} \left(\widetilde{v}_i^2 \langle \widetilde{\Phi}_i, \widetilde{\Phi}_i \rangle \right)$$

and generally for all $m \geq 0$

$$\nu^{(m)}(v) := \int_{\operatorname{spt} \xi} v(\xi)^m \mu(d\xi) \tag{9}$$
$$= \sum_{i_1=1}^{\widetilde{P}} \cdots \sum_{i_m=1}^{\widetilde{P}} \left(\widetilde{v}_{i_1} \cdots \widetilde{v}_{i_m} \langle \widetilde{\Phi}_{i_1}, \cdots \widetilde{\Phi}_{i_m} \rangle \right).$$

The integrals can be evaluated using Gauss Quadrature. The calculation of the higher moments ($m > 2$) can be a computationally formidable task due to the large number of integrals that need to be evaluated. However, the majority of integrals are zero due to (power) orthogonality.

4. EXPERIMENTAL DESIGN FOR MODEL DISCRIMINATION

Optimal experimental design for the discrimination of two models aims to determine an input sequence that enables associating the actual system measurements $\widehat{y}$ with the outputs of either model $m^{[1]}$ or model $m^{[2]}$, which requires that the PDFs of the outputs of the two models be separated in the presence of probabilistic uncertainties. The probability of misclassification (i.e., choosing the wrong model) is directly related to the similarity of two PDFs (Anderson, 2003): the more similar the distributions are, the larger the error. Hence, the probability of misclassification (i.e., the Bayes error) should be minimized to reduce the overlap of the output PDFs of the two models.

Next, a measure for the dissimilarity of two PDFs is presented and its relationship to the Bayes error is established. PC will then be used to propagate the probabilistic uncertainties of the initial conditions and parameters to approximate the PDFs of the outputs. One of the challenges is to include the measurement noise into the analysis. Finally, a nonlinear optimization problem is formulated to design the input sequence for model discrimination.

4.1 Measure for Dissimilarity of Probability Densities

A measure closely related to the Bayes error is the Bhattacharyya coefficient Bhattacharya and Toussaint (1982); Kailath (1967). The Bhattacharyya coefficient is defined for a two-hypotheses (see (Bhattacharya and Toussaint, 1982) for multiple hypotheses) problem by

$$\rho^{[1,2]} := \int_{-\infty}^{\infty} \sqrt{\mu^{[1]}(\xi)\mu^{[2]}(\xi)} d\xi, \tag{10}$$

where $\mu^{[1]}$ and $\mu^{[2]}$ are two continuous PDFs corresponding to the two hypotheses. The Bhattacharyya coefficient $\rho^{[1,2]}$ will be larger if the similarity of the PDFs is larger. The coefficient is one for identical distributions and is zero if the distributions do not overlap. The advantage of the Bhattacharyya coefficient is its validity for any distribution and, therefore, its suitability as a generic model discrimination criterion. In addition, the Bhattacharyya coefficient gives an upper bound on the Bayes probability of misclassification $P_e^{[1,2]}$. When the prior probabilities of the two model hypotheses are equal, the Bayes probability of misclassification is given by (Bhattacharya and Toussaint, 1982)

$$P_e^{[1,2]} \leq \frac{1}{2} \rho^{[1,2]}. \tag{11}$$

In this work, the Bhattacharyya coefficient is used for characterizing the similarity of the output PDFs of the two models. Expression (10) merely allows the comparison of univariate PDFs (i.e., one output at one measurement time point). However, the formulation of the model discrimination problem requires taking into account multiple outputs and multiple measurement time points, which motivates the use of the criterion

$$\Lambda^{[1,2]} := \prod_{k=1}^{m_y} \prod_{l=1}^{n_y} \rho_{lk}^{[1,2]}, \tag{12}$$

where $\rho_{lk}^{[1,2]}$ denotes the Bhattacharyya coefficient of the l^{th} output at time instant k of models $m^{[1]}$ and $m^{[2]}$. $\Lambda^{[1,2]}$ corresponds to the product of the different Bhattacharyya coefficients for all outputs and at all measurement time points. If $\Lambda^{[1,2]} = 0$, then the PDFs do not overlap at least at one time point and, therefore, the misclassification probability is zero if the corresponding output is measured at the corresponding time point. When $\Lambda^{[1,2]} > 0$, however, the PDFs overlap at all time points. In this case, the upper bound on the misclassification error for the different outputs at the different time points can be determined using (11).

4.2 Moment-Based Approximation of Output Probability Densities Corrupted by Noise

According to (2), the measurement noise affects the model outputs multiplicatively. The PDF of the noise-free part $y^{[j]} = h^{[j]}(x^{[j]}, p^{[j]}, u)$ of the model output can be determined using PC-based Monte Carlo simulations (Mesbah et al., 2014a), or by approximating the PDFs using the moments (see (9)). This paper employs the latter approach.

Provided that the measurement noise $w_{lk}^{[j]}$ is independent from $y_{lk}^{[j]}$ in (2), the known moments $\nu^{(i)}(w_{lk}^{[j]})$ and the moments $\nu^{(i)}(\widetilde{y}_{lk}^{[j]})$ obtained using PC expansions can be utilized to determine the m^{th} moments of the PDF of the noisy outputs $\mu(\widehat{y}_{lk})$:

$$\nu^{(m)}(\widehat{y}_{lk}^{[j]}) = \sum_{i=0}^{m} \binom{m}{i} \nu^{(i)}(\widetilde{y}_{lk}^{[j]})\nu^{(m-i)}(\widetilde{y}_{lk}^{[j]})\nu^{(m-i)}(w_{lk}^{[j]}). \tag{13}$$

Expression (13) follows directly from the definition of the expectation and (raw) moments, as well as the property that $y_{lk}^{[j]}$ and $w_{lk}^{[j]}$ are stochastically independent.

To compute $\Lambda^{[1,2]}$, an approximation of the PDF in form of an analytic expression is derived using the moments. Different approaches to determine the PDFs based on moments have been proposed in the literature. One approximation is given by the product of a PDF of a chosen base density function (such as the Normal or Uniform distributions) and a polynomial (Provost, 2005). The coefficients of the polynomial can be explicitly determined from the moments of the base density and the PDF that is to be approximated. This approach results in a closed-form solution that is easy to implement. However, a large number of moments is required to obtain a good approximation and the approximation accuracy depends on the congruence of the tails of the base density and the to-be-approximated PDF. Another approach to approximate the PDF is maximum entropy estimation using a Gaussian Mixture Model, along with constraints on the moments (Dutta and Bhattacharya, 2010). This approach is very flexible and is suitable to approximate multimodal distributions. However, besides an appropriate choice of the number of Gaussian bases, the approach requires solution of a nonlinear optimization problem, which would further complicate the experimental design problem.

In this work, a simple method is used to approximate the output PDFs by employing the four-parameter Beta distribution (Hanson, 1991). The four-parameter Beta distribution is a family of continuous probability distributions parameterized by two shape parameters ($\alpha > 0$, $\beta > 0$) and two parameters defining the bounded support (lb and ub for lower and upper bound). The distribution can be defined in terms of the gamma function Γ by

$$\beta_{\text{PDF}}(\zeta, \alpha, \beta, lb, ub) := \frac{(\zeta - lb)^{\alpha-1}(ub-\zeta)^{\beta-1}}{(ub-lb)^{\alpha+\beta-1}} \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}.$$

All four parameters in this expression can be estimated based on the method of moments (Hanson, 1991) using the first four central moments (mean, variance, skewness, and excess kurtosis). The advantage of this approach is that it is flexible and an analytic form of the PDF can be easily obtained. In addition, the resulting β_{PDF} distributions possess the same first four moments as the noisy output PDFs as given by (13) if these are used to determine the four parameters. Since the calculation of the moments based on the coefficients of the PC expansions is computationally demanding for higher order moments (see Eqs. (9)), having an analytic form of a distribution with a small number of moments provides a reasonable tradeoff between the accuracy of PDF construction from moments and the computational efficiency. Furthermore, a closed-form solution for the Bhattacharyya coefficient (10) is available for the Beta distribution, which means that (12) can be easily obtained.

In the sequel, $\widehat{\Lambda}_{\beta}^{[1,2]}$ denotes the approximation of the model discrimination criterion in (12), which is obtained using the four-parameter Beta distribution β_{PDF}.

4.3 Least-costly Input Design for Model Discrimination

In the so-called least-costly experimental design framework (Gevers, 2005), some norm of the input signal is often minimized. In this work, the input design problem for model discrimination is formulated to minimize the function $J\big(\bar{u}(t_{u,1}), \ldots, \bar{u}(t_{u,m_u})\big) := \sum_{k=1}^{m_u} \|\bar{u}(t_{u,k})\|_2$, while ensuring that the outputs of competing models are separated in the presence of uncertainties. Based on the above results, the input design problem can be cast as

$$\begin{aligned}
\underset{\bar{u}(t_{u,1}),\ldots,\bar{u}(t_{u,m_u})}{\text{infimum}} \quad & J\big(\bar{u}(t_{u,1}), \ldots, \bar{u}(t_{u,m_u})\big) \\
\text{subject to:} \quad & \dot{\widetilde{x}}^{[1]} = \widetilde{f}^{[1]}(\widetilde{x}^{[1]}, \widetilde{p}^{[1]}, u, \widetilde{\Phi}^{[1]}), \quad \widetilde{x}^{[1]}(0) \\
& \widetilde{y}^{[1]} = \widetilde{h}^{[1]}(\widetilde{x}^{[1]}, \widetilde{p}^{[1]}, u, \widetilde{\Phi}^{[1]}) \\
& \dot{\widetilde{x}}^{[2]} = \widetilde{f}^{[2]}(\widetilde{x}^{[2]}, \widetilde{p}^{[2]}, u, \widetilde{\Phi}^{[2]}), \quad \widetilde{x}^{[2]}(0) \\
& \widetilde{y}^{[2]} = \widetilde{h}^{[2]}(\widetilde{x}^{[2]}, \widetilde{p}^{[2]}, u, \widetilde{\Phi}^{[2]}) \\
& \widehat{\Lambda}_{\beta}^{[1,2]} \le \overline{\Lambda} \\
& \bar{u}(t_{u,k}) \in \mathcal{U}_k, \quad \forall k = 1, \ldots, m_u,
\end{aligned} \tag{14}$$

where $\overline{\Lambda}$ is some nonnegative threshold that provides an upper bound on the worst-case misclassification error. With (11) and (12), $\overline{\Lambda}$ can be defined by

$$\overline{\Lambda}^{\frac{1}{m_y n_y}} \le 2P_{e,\max}, \tag{15}$$

where $P_{e,\max}$ is the maximum admissible misclassification error for a measured output at a certain time instant.

5. NUMERICAL ILLUSTRATION

This section demonstrates the proposed optimal input design approach for model discrimination.

5.1 Model Descriptions

Consider the Michaelis-Menten and Henri mechanisms as two model hypotheses for an enzyme-catalyzed reaction (see Henri (2006); Rumschinski et al. (2010)). In both biochemical reaction models, an enzyme E and a substrate S form an enzyme-substrate complex C, which is converted to a final product P. The models have two conservation relations in which the total enzyme and substrate concentrations are assumed to be one. The dynamic models are described according to the law of mass action, where x_1 and x_2 denote the substrate and complex, respectively. The model for the Henri-mechanism is described by

$$m^{[1]}: \begin{cases} \dot{x}_1^{[1]} = (p_1^{[1]} + p_3^{[1]})(x_2^{[1]} - 1)x_1^{[1]} + (p_2^{[1]} + u)x_2^{[1]} \\ \dot{x}_2^{[1]} = p_1^{[1]}(1 - x_2^{[1]})x_1^{[1]} - (p_2^{[1]} + u)x_2^{[1]}. \end{cases}$$

The model for the Michaelis-Menten mechanism is

$$m^{[2]}: \begin{cases} \dot{x}_1^{[2]} = p_1^{[2]}(x_2^{[2]} - 1)x_1^{[2]} + (p_2^{[2]} + u)x_2^{[2]} \\ \dot{x}_2^{[2]} = p_1^{[2]}(1 - x_2^{[2]})x_1^{[2]} - ((p_2^{[2]} + u) + p_3^{[2]})x_2^{[2]}. \end{cases}$$

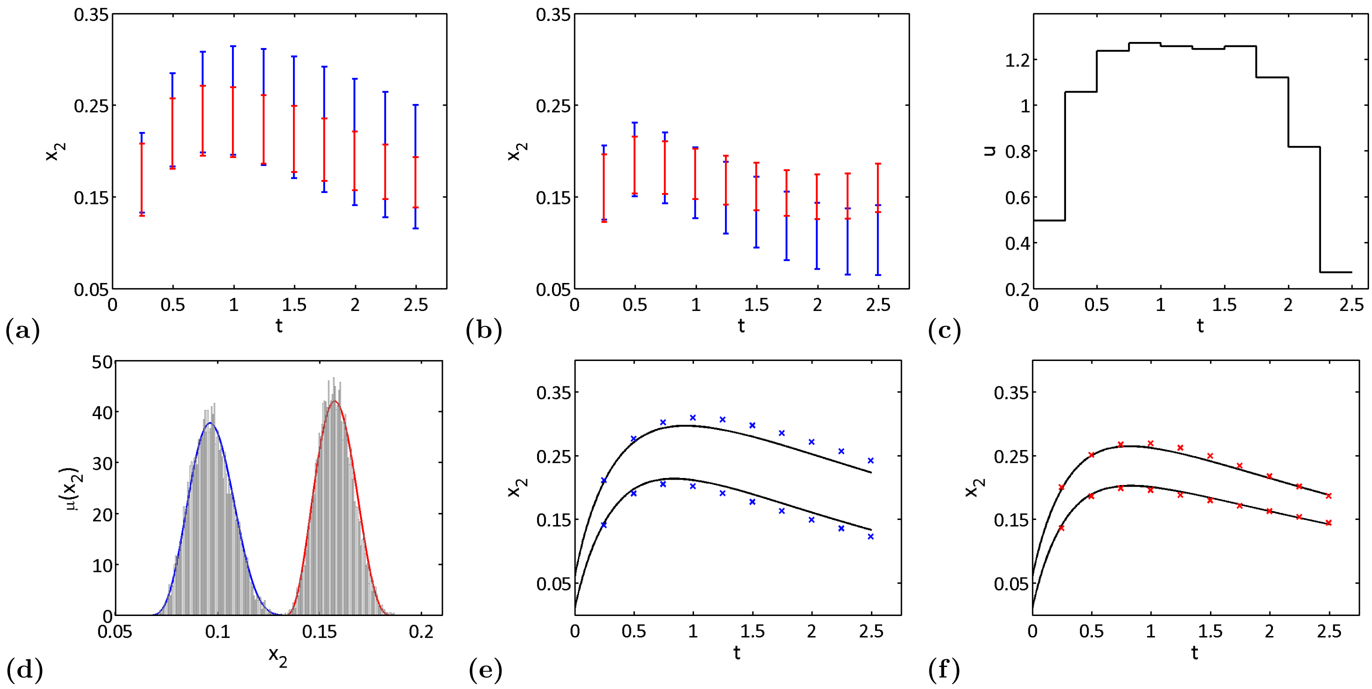

Fig. 1. Illustration of the discrimination of models $m^{[1]}$ (blue) and $m^{[2]}$ (red) by input design. (a) The support of the output PDFs for an initial input sequence. (b) The support of the output PDFs when the designed input sequence is applied. (c) The designed input sequence. (d) Output PDFs at time instant $t = 2.5$ as obtained for the applied input and comparison with Monte-Carlo simulations (black thin lines show the histogram for the Monte-Carlo simulations). (e) and (f) show the comparison of minimal/maximal values from the Monte-Carlo simulations (black lines) with the support predicted by the Beta distribution fits for models $m^{[1]}$ and $m^{[2]}$.

The initial conditions $x_1(0)$ and $x_2(0)$ are uniformly distributed on the intervals $[0.93, 0.98]$ and $[0.01, 0.06]$, respectively. The parameters p_1, p_2, p_3 are uniformly distributed on $[0.9, 1.1]$. The input u is assumed to affect the reaction parameter p_2 in an additive manner.

PC expansions of order $P = 2$ were used to approximate the two models. Fig. 1(a) shows the measured output $y = x_2$. As can be seen, the support of the probability densities overlap, which indicates that the models cannot be discriminated adequately even when the output was not corrupted by measurement noise. Thus, an input sequence is designed to discriminate the models.

5.2 Input Design

The experimental design was performed for a time horizon of 2.5 min, which was split into 10 equidistant time intervals. The input sequence was discretized in a piecewise manner. The input was constrained by $0 \leq u_k \leq 5$, $k = 1, \ldots, 10$. Output measurements were taken at $t_{y,1} = 0, t_{y,2} = 0.25, \ldots, t_{y,11} = 2.5$ min. The measurements were corrupted by a white noise sequence, which has a normal distribution with zero mean and standard deviation 0.0326.

The nonlinear optimization problem (14) with $P_{e,\max} = 0.174$ was solved using the Matlab routine `fmincon`, in which the initial input profile was defined as $u_0 = [0, 1, 1, 0, \ldots, 0]$. The optimization was terminated after 15 iterations, which took all together about 60 min. Fig. 1(b) indicates that the designed input sequence (shown in Fig. 1(c)) enabled an adequate separation of the noisy outputs of the two models and, therefore, facilitates the task of model discrimination.

In Figs. 1(d)–(f), 10,000 Monte-Carlo simulations are compared with the PC simulation and subsequent Beta distribution fit. Fig. 1(d) shows the PDF of the noisy model outputs at time instant 2.5 min, which demonstrates that the output PDFs are separated and that the proposed PC approach agrees very well with Monte-Carlo simulations. Figs. 1(e)–(f) compare Monte-Carlo simulations with PCE simulation, which indicate a satisfactory prediction of the support of the output PDFs at different time instants by the PC expansions.

6. CONCLUSION

This paper presents a method for input design for optimal model discrimination. Uncertainties of the parameters and initial conditions are taken into account using a Polynomial Chaos approach to approximate the PDF of the outputs. An empirical measure for model discrimination based on the Bhattacharyya coefficient is proposed that is especially suited for the proposed nonlinear optimization problem. The measure requires the knowledge of the probability densities of the noisy output. A simple moment-based approximation of the PDF of the noisy output was employed using the four-parameter Beta distribution.

The four-parameter Beta-distribution is a very versatile distribution that can approximate many other distributions including the Normal, Uniform, or Log-normal distributions. However, the approximation can be unsatisfactory, e.g., for multimodal distributions. The approximation quality can be tested by PCE simulation and evaluating the fits visually or quantitatively. In case of low approximation quality, other approximation techniques should be used as described in the text.

ACKNOWLEDGMENT

We thank R. Bhushan Gopaluni for helpful discussions.

REFERENCES

Anderson, T.W. (2003). *An Introduction to Multivariate Statistical Analysis*. John Wiley, New York.

Atkinson, A.C. and Fedorov, V. (1975). Optimal design: Experiments for discriminating between several models. *Biometrika*, 62(2), 289–303.

Bhattacharya, B.K. and Toussaint, G.T. (1982). An upper bound on the probability of misclassification in terms of Matusita's measure of affinity. *Annals of the Institute of Statistical Mathematics*, 34(1), 161–165.

Blanke, M., Kinnaert, M., Lunze, J., and Staroswiecki, M. (2006). *Diagnosis and Fault-Tolerant Control*. Springer, Heidelberg, Berlin.

Borchers, S., Raković, S., and Findeisen, R. (2011). Set membership parameter estimation and design of experiments using homothety. In *18th IFAC World Congress*, 9034–9040. Milan, Italy.

Box, G.E.P. and Hill, W.J. (1967). Discrimination among mechanistic models. *Technometrics*, 9, 57–71.

Campbell, S.L. and Nikoukhah, R. (2004). *Auxilary Signal Design for Failure Detection*. Princeton University Press, New Jersey.

Chen, B.H. and Asprey, S.P. (2003). On the design of optimally informative dynamic experiments for model discrimination in multiresponse nonlinear situations. *Industrial & Engineering Chemistry*, 42, 1379–1390.

Dutta, P. and Bhattacharya, R. (2010). Nonlinear estimation of hypersonic state trajectories in bayesian framework with polynomial chaos. *Journal of Guidance, Control, and Dynamics*, 33(6), 1765–1778.

Fedorov, V.V. (1972). *Theory of Optimal Experiments*. Academic Press, New York.

Fisher, J. and Bhattacharya, R. (2009). Linear quadratic regulation of systems with stochastic parameter uncertainties. *Automatica*, 45(12), 2831–2841.

Gautschi, W., Friedman, R.S., Burns, J., Darjee, R., and Mcintosh, A. (2004). *Orthogonal Polynomials: Computation and Approximation, Numerical Mathematics and Scientific Computation Series*. Oxford University Press, Oxford, U.K.

Georgiev, D. and Klavins, E. (2008). Model discrimination of polynomial systems via stochastic inputs. In *Proc. of the IEEE Conference on Decision and Control*, 3323–3329. Cancun, Mexico.

Gerritsma, M., Van der Steen, J.B., Vos, P., and Karniadakis, G. (2010). Time-dependent generalized polynomial chaos. *Journal of Computational Physics*, 229(22), 8333–8363.

Gevers, M. (2005). Identification for control: From the early achievements to the revival of experiment design. *European Journal of Control*, 11, 1–18.

Goodwin, G.C. and Payne, R.L. (1977). *Dynamic System Identification: Experiment Design and Data Analysis*. Academic Press, New York.

Halder, A. and Bhattacharya, R. (2012). Further results on probabilistic model validation in Wasserstein metric. In *Proc. of the IEEE Conference on Decision and Control*, 5542–5547.

Hanson, B.A. (1991). *Method of Moments Estimates for the Four-parameter Beta Compound Binomial Model and the Calculation of Classification Consistency Indexes*, volume 91. American College Testing Program.

Henri, V. (2006). General theory of the action of some glycoside hydrolases. *Comptes Rendus Biologies*, 329(1), 47–50.

Kailath, T. (1967). The divergence and Bhattacharyya distance measures in signal selection. *IEEE Transactions on Communication Technology*, 15(1), 52–60.

Kim, K.K.K., Shen, D.E., Nagy, Z.K., and Braatz, R.D. (2013). Wiener's polynomial chaos for the analysis and control of nonlinear dynamical systems with probabilistic uncertainties. *IEEE Control Systems Magazine*, 33(5), 58–67.

Mesbah, A., Streif, S., Findeisen, R., and Braatz, R.D. (2014a). Active fault diagnosis for nonlinear systems with probabilistic uncertainties. In *Proc. of the 19th IFAC World Congress*. Cape Town, South Africa. In press.

Mesbah, A., Streif, S., Findeisen, R., and Braatz, R.D. (2014b). Stochastic nonlinear model predictive control with probabilistic constraints. In *Proc. of the American Control Conference*. Portland, Oregon. In press.

Milovanović, G.V. (2001). Quadratures with multiple nodes, power orthogonality, and moment-preserving spline approximation. *Journal of Computational and Applied Mathematics*, 127(1), 267–286.

Ohtsuka, T. (2005). Model structure simplification of nonlinear systems via immersion. *IEEE Transactions on Automatic Control*, 50(5), 607–618.

Oladyshkin, S. and Nowak, W. (2012). Data-driven uncertainty quantification using the arbitrary polynomial chaos expansion. *Reliability Engineering & System Safety*, 106, 179–190.

Paulson, J.A., Raimondo, D.M., Braatz, R.D., Findeisen, R., and Streif, S. (2014). Active fault diagnosis for nonlinear uncertain systems. In *Proc. of the European Control Conference*. Strasbourg, France. In press.

Provost, S.B. (2005). Moment-based density approximants. *The Mathematica Journal*, 9(4), 727–756.

Pukelsheim, F. and Rosenberger, J.L. (1993). Experimental designs for model discrimination. *Journal of the American Statistical Association*, 88(422), 642–649.

Rumschinski, P., Borchers, S., Bosio, S., Weismantel, R., and Findeisen, R. (2010). Set-base dynamical parameter estimation and model invalidation for biochemical reaction networks. *BMC Systems Biology*, 4, 69.

Scott, J.K., Findeisen, R., Braatz, R.D., and Raimondo, D.M. (2013). Design of active inputs for set based fault diagnosis. In *Proc. of the American Control Conference*, 3567–3572. Washington D.C.

Walter, E. and Pronzato, L. (1990). Qualitative and quantitative experiment design for phenomenological models–A survey. *Automatica*, 26(2), 195–213.

Wiener, N. (1938). The homogeneous chaos. *American Journal of Mathematics*, 60, 897–936.

Xiu, D. and Karniadakis, G.E. (2002). The Wiener-Askey polynomial chaos for stochastic differential equations. *SIAM Journal of Scientific Computation*, 24, 619–644.

Zarrop, M.B. (1979). *Optimal Experiment Design for Dynamic System Identification*. Springer-Verlag, New York.