

Model Discrimination of Polynomial Systems via Stochastic Inputs

D. Georgiev and E. Klavins

Abstract— Systems biologists are often faced with competing models for a given experimental system. Unfortunately, performing experiments can be time-consuming and expensive. Therefore, a method for designing experiments that, with high probability, discriminate between competing models is desired. In particular, biologists often employ models comprised of polynomial ordinary differential equations that arise from biochemical networks. Within this setting, the discrimination problem is cast as a finite-horizon, dynamic, zero-sum game in which parameter uncertainties in the model oppose the effort of the experimental conditions. The resulting problem, including some of its known relaxations, is intractable in general. Here, a new scalable relaxation method that yields sufficient conditions for discrimination is developed. If the conditions are met, the method also computes the associated random experiment that can discriminate between competing models with high probability, regardless of the actual system behavior. The method is illustrated on a biochemical network with an unknown structure.

I. INTRODUCTION

Systems biologists are often faced with competing dynamical models of experimental systems. For example, a newly discovered regulatory protein identified as an inhibitor may act on one of several possible genes in a pathway, leading to different possible biochemical network models. One way to distinguish between different models is to create knockout cell-lines and examine the steady-state behavior of the resulting mutants. However, such experiments are tedious and may be quite difficult depending on how vital the considered network is to the normal operation of the cell. An alternative approach is to run dynamical experiments on cells and interpret their transient response to distinguish between competing models. For example, a nutrient or a chemical signal can be changed in a time-varying manner (i.e. it is an input signal to the system) and the intensity of a fluorescent marker incorporated into the network can be observed, as was recently demonstrated with osmotic pressure regulation in yeast [1].

Dynamical experiments are still expensive to set up, but they are potentially less invasive and much more informative than static experiments. Within this setting, we address the question: What experiments should be performed on the physical system to ensure that as many candidate models are invalidated by the experimental results as possible? In particular: given a set of candidate models, can we define a probability distribution, which we call a *disparity certificate*,

over the possible input signals that maximally distinguishes the candidate models? If no input signals distinguish the candidate models, the experiments are not worth doing until better candidates are derived. If the candidate models are distinguishable, probing the actual experimental system with the resulting disparity certificate must invalidate at least one of the models (see Figure I).

Invalidation is made difficult by uncertainty in the experimental system. In systems biology, uncertainty arises particularly from unknown and even time varying reaction rates (e.g. some functions slow down during cell division). Thus, an erroneous model may not be invalidated by an experiment because the uncertain parts of the physical system ‘conspire’ against the researcher to produce outputs that seem consistent with the model. In our search for a disparity certificate we must expect that uncertainty in the system will work to make candidate models indistinguishable. The resulting problem becomes a game: the input signal tries to force the outputs of two candidate models to be different while the uncertainty tries to make the outputs the same.

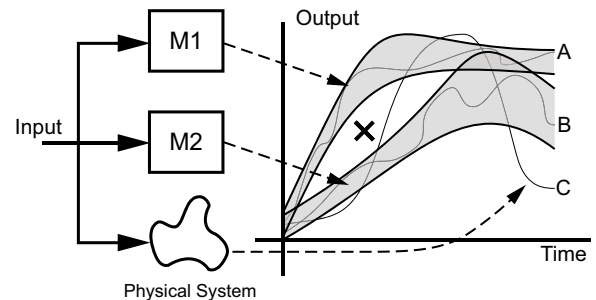


Fig. 1. Model Discrimination for experimental design. X represents the disparity certificate. Shaded regions represent all possible trajectories corresponding to the input u . Trajectories A, B, C are possible outputs of the physical system. For output A model 2 is invalidated, for output B, model 1 is invalidated and for output C both models are invalidated.

In this paper, we suppose we are given two different candidate models that have the same input and output spaces. We put uncertainty in the initial conditions and parameters of the models. We pose two problems: *Model Discrimination* and *Model Invalidation*. The first problem we set up as a *maximin* problem (maximizing over all input signal distributions and minimizing over all disturbances the difference between the outputs of the two models). Since the resulting optimization problem is intractable, we introduce a relaxation (based on the theory of moments [2]) that we can solve more efficiently. The second problem is then easily solved using the disparity certificate produced by the first problem. We have implemented the method in *MATLAB* and demonstrate

This work was supported in part by the following grants: NIH/NIDCR, #T32 DE007132-26 and ARO MURI, #W 911 NF 0710287.

D. Georgiev and E. Klavins are with the Department of Electrical Engineering, University of Washington, Seattle, WA 98195-2500, USA dgeorgie@u.washington.edu, klavins@u.washington.edu

it on two problems involving candidate parameter regimes in biochemical networks.

II. RELATED WORK

Model discrimination is widely studied in the literature. Experimental design based on model discrimination has also received a considerable attention [3–7]. The traditional approach is to apply Bayesian conditioning and discriminate models using criterion such as the maximum likelihood [3, 8, 9]. Methods based on the Bayesian approach assume that the real world system behaves according to a given candidate model with a known probability. As a result, the success of these methods depends on the quality of prior information. More recent methods are based on deterministic models [4–6], as is the case in this paper. In [6], a method is developed that computes the initial state to maximize an upper bound on the distance between outputs of competing models. The models, however, are deterministic and, by maximizing the upper bound, the method cannot guarantee discrimination. In [4] algorithms for efficiently discriminating models from experimental data are developed, however, no inputs are considered. Finally, a great deal of literature on auxiliary input design for fault detection deals with model discrimination [8, 10]. The majority of this work is based on linear models and Bayesian conditioning. In our work, nonlinear (polynomial) systems of difference equations with unstructured uncertainty are considered and input signal distributions guaranteeing discrimination are computed.

III. PROBLEM STATEMENT

A. Informal Description

The work herein is motivated by biochemical experimental design. The following is an informal description of our work; the formal description is in the next section. We consider models comprising polynomial difference equations and linear outputs. An experiments controlled variables are modeled as control inputs to the system. We assume feedback is not possible or practical during the execution of the experiment and, therefore, restrict the control inputs to be functions of only time. An experiments uncontrolled variables are modeled as disturbances. Disturbances are unstructured, i.e., can be effected by past events in an arbitrary way, and the only known information about them is their domains. Note that disturbance information is part of the model. The methods herein do not require that the actual experiments uncontrolled variables have the same domain as the disturbances.

Computation in experiments is performed in two steps: 1) experimental design and 2) data analysis. The two steps are matched by the two problems considered in this paper. The first problem considers feasibility of model discrimination and the computation of discriminating inputs, called disparity certificates.

Model Discrimination Problem (MDP) (informal). Given a pair of candidate models with the same input and output spaces, find an input, called the disparity certificate, that yields different outputs for all possible disturbances.

If a disparity certificate exists, we can hope to implement it in the experiment and learn which model does not represent the system. Identifying the inconsistent model is called model invalidation. The second problem seeks to invalidate the candidate models based on experimental data from a series of experiments.

Model Invalidation Problem (MIP) (informal). Given the inputs and outputs for a series of executed experiments, find which candidate model maps the inputs to different outputs for all possible disturbances.

Model invalidation is studied in [11]. As will be shown, disparity certificates enable a model to be invalidated by simultaneously comparing it with experimental data from multiple trials. The method in [11] does not consider such comparisons and the extension is not clear. We address the problem within the framework of this paper and show that if we can find a disparity certificate, we can invalidate a model.

B. Formal Description

The following notation is used. Lowercase letters are used to denote functions and vectors. A sequence of vectors $\{r(1), \dots, r(K)\}$ is written as r^K . Uppercase letters are used to denote matrices and random variables. As with a sequence of vectors, a sequence of random variables $\{R(1), \dots, R(K)\}$ is written as R^K . For any discrete random variable R , $\mathbb{P}(r)$ is the probability that $R = r$ (the random variable associated with \mathbb{R} will be clear from the context). For any sequence of random variables R^K , $\mathbb{P}(r(K) | r^{K-1})$ is the probability that $R(K) = r(K)$ given that $R^{K-1} = r^{K-1}$. Uppercase script letters denote sets. For a given set \mathcal{R} , \mathcal{R}^K is the cartesian product $\times_{i=1}^K \mathcal{R}$. The set of all probability distributions over \mathcal{R} is denoted by $\Delta(\mathcal{R})$. For a vector $\beta \in \mathbb{N}^K$, we use the notation from [2] and write the monomial $r_1^{\beta_1} \dots r_K^{\beta_K}$ as r^β .

Throughout this paper, we consider a pair of models described by discrete time difference equations of the form

$$M_i = \begin{cases} x_i(k+1) & = f_i(x_i(k), w_i(k), u(k)), \\ y_i(k) & = C_i x_i(k), \end{cases} \quad (1)$$

where $k \in \{1, \dots, K\}$ and $i \in \{1, 2\}$. The initial state satisfies $x_i(1) \in \mathcal{X}_i(1) \subset \mathbb{R}^{n_i}$, the control input satisfies $u(k) \in \mathcal{U} \subset \mathbb{R}^m$, and the disturbance satisfies $w_i(k) \in \mathcal{W}_i \subset \mathbb{R}^{d_i}$. Throughout the paper, we concatenate a pair of models to form a composite model M with state $x = (x_1, x_2) \in \mathbb{R}^n$, state transition function $f = (f_1, f_2)$, control input u , disturbance $w = (w_1, w_2) \in \mathbb{R}^d$, and output $y = y_1 - y_2$.

We make the following assumptions.

Assumption

A1. State transition function f is a polynomial function with respect x and w , i.e.,

$$f(x, w, u) = \sum_{\alpha} a_{\alpha}(u) (x, w)^{\alpha}. \quad (2)$$

A2. The set of control inputs \mathcal{U} is finite.

A3. The sets $\mathcal{X}(1)$ and \mathcal{W} are compact intervals in their respective Euclidean spaces.

Before we proceed to address the issue of model discrimination, we must define how the values of the control inputs and the disturbances are selected. Experimental data is often evaluated after the completion of an experiment. Hence, we require that open loop control input be selected. Open loop evaluation, however, does not imply that the input is selected deterministically. Indeed, random control inputs are commonly used in system identification to provide persistent excitation [12], and in game theory it is well known that random decisions often outperform deterministic decisions [13]. In this work we consider random control inputs, and, through example, show that this increases the class of systems that can be discriminated.

Due to physical, experimental, or computational constraints, not all possible trajectories in \mathcal{U}^K may be allowed. Define the set of control inputs allowed to follow u^{k-1} by $\mathcal{S}_k(u^{k-1}) \subseteq \mathcal{U}$. The set of allowable control input trajectories of length k is denoted by \mathcal{T}_k and is defined recursively by $\mathcal{T}_i = \{u^i \in \mathcal{U}^i | u^{i-1} \in \mathcal{T}_{i-1}, u(i) \in \mathcal{S}_i(u^{i-1})\}$. Therefore, a control input is a random sequence U^K with the sample space \mathcal{T}_K .

Whether an input discriminates two models depends on how the disturbance is realized. The disturbance may be unknown but fixed a priori. This would imply that no matter what the input is, the uncertainty remains unchanged. If the inputs do affect the disturbance, the effect may be causal. This would imply that the uncertainty can change if any of the past inputs are changed. If the inputs affect the uncertain parts in a noncausal manner, the uncertainty can change if any inputs (past or future) are changed. Which of the three alternatives is chosen depends on the system. However, in biological systems, the uncertainty (e.g., the reaction rates) is assumed to depend on the current state of the system and is therefore causal. Moreover, viable experiments are duplicable. Therefore, we assume that the disturbance at any time instance is a function of all past exogenous inputs to the system and this function is the same for different experimental trials. Note that the initial state is treated as part of the disturbance, which leaves the control inputs as the only exogenous inputs. The value of the disturbance given that the control inputs up to time $k-1$ are given by u^{k-1} is denoted by $w(k; u^{k-1})$. Similarly, $x(k; u^{k-1})$ and $y(k; u^{k-1})$ denote the state and output, respectively, that result from the control inputs u^{k-1} and the disturbances $w(i; u^{i-1}), i \in \{1, \dots, k\}$. A random control input U^{k-1} yields the random disturbance $w(k; U^{k-1})$, the random state $x(k; U^{k-1})$, and the random output $y(k; U^{k-1})$.

The disparity of a pair of models is defined in terms of a disparity certificate.

Definition 1 (Disparity Certificate). A pair of models (M_1, M_2) is said to have a *disparity certificate* U^K if, for all conditional disturbances $w(k; u^{k-1}) \in \mathcal{W}, k \in \{1, \dots, K\}$, the output is nonzero with positive probability, i.e., for some

$k \in \{1, \dots, K\}$,

$$\mathbb{P}(\|y(k; U^{k-1})\|_2 > 0) > 0.$$

The two problems considered in this paper can now be formally defined. The model discrimination problem is cast as a dynamic game where the disturbance uses information about all past control inputs to try and keep the outputs of the two models equal.

Model Discrimination Problem (MDP). Given a composite model M , find a U^* with the distribution P^* that solves the following maximin problem:

$$\max_P \min_w \sum_{u^K \in \mathcal{T}_K} \mathbb{P}(u^K) \sum_{k=1}^{K+1} \|y(k; u^{k-1})\|_2,$$

subject to

$$\begin{aligned} x(k+1; u^k) &= f(x(k; u^{k-1}), w(k; u^{k-1}), u(k)) \\ y(k; u^{k-1}) &= Cx(k; u^{k-1}). \end{aligned}$$

Note that the cost in MDP is just the expected value with respect to the control input U^* . A disparity certificate exists if and only if the optimal cost in MDP is positive. Note that the polynomial constraints make MDP nonconvex in general.

Next we address the model invalidation problem. Suppose that ℓ trials of a particular experiment have been performed. We would like to know if one of the models is inconsistent, and thereby invalidated, by the trials. Recall that the conditional disturbances are assumed to be the same in different trials. Therefore, if a model is consistent, it should be able to approximate outputs from all trials at the same time.

Model Invalidation Problem (MIP). Let the results of ℓ experimental trials be given by the set of input-output pairs $\{(u_{e,1}^K, y_{e,1}^K), \dots, (u_{e,\ell}^K, y_{e,\ell}^K)\}$. For a given model M_i , the *model invalidation problem* is defined as follows.

$$\minimize_w \sum_{j=1}^{\ell} \sum_{k=1}^{K+1} \|y_i(k; u_{e,j}^{k-1}) - y_{e,j}(k)\|_2^2,$$

subject to

$$\begin{aligned} x_i(k+1; u_{e,j}^k) &= f_i(x_i(k; u_{e,j}^{k-1}), w_i(k; u_{e,j}^{k-1}), u_{e,j}(k)) \\ y_i(k; u_{e,j}^{k-1}) &= C_i x_i(k; u_{e,j}^{k-1}). \end{aligned}$$

If the optimal cost of MIP is positive for model M_i , then the model is invalidated. Notice that MIP also presents a nonconvex optimization problem.

The following result follows directly from Definition 1, MDP, and MIP.

Proposition III.1. *Let a pair of models (M_1, M_2) have a disparity certificate U^* . Let $|SUP(U^*)|$ be the cardinality of the sample space of U^* . Then at most $|SUP(U^*)|$ different experiments are required to invalidate at least one model.*

Therefore, the existence of a disparity certificate implies that, given enough experimental trials, useful information (i.e., invalidation of a model) must eventually result. Note that experiments are typically expensive and difficult to

organize but executing multiple parallel trials is relatively easy.

IV. SOLUTION

To discriminate two models, one must solve MDP. First, for any control input U , the corresponding global minimizer (the worst disturbance) w^* must be found. However, finding a global minimizer in MDP is in general numerically intractable due to the polynomial state transition equality constraints. Second, the maximizing control input must be found. The first part of this section develops a method for computing the lower bound on the optimal cost for a given control input. The second part of the section appends this method to allow for simultaneous computation of the lower bound maximizing control input.

A. Worst Disturbance

Several methods exist for minimizing the cost in MDP. The brute force method is to discretize the state space and numerically solve the Bellman equation, see [14] for examples. The state space dimension of the class of systems considered herein prohibits such an approach. Other methods yield a lower bound on the cost. One such approach is to use difference inclusions to compute the set of reachable states and then minimize the cost over all trajectories [15]. Difference inclusions however are overly conservative in that future states depend only on the current set of reachable states and not on the actual current state value. Another approach is to approximate the global optimal solution using sum-of-squares convex relaxation [2, 16]. This approach yields a lower bound on the optimal solution up to any specified tolerance. However, it is only numerically feasible for problems comprising relatively few optimization variables; the method presented in [2] applied to MDP with 10 time steps, and 5 dimensional state space and 5 dimensional disturbance space at each time step requires approximately 100 million optimization variables for a first order relaxation.

In this section, we construct a method for computing the lower bound on the optimal cost using the theory of moments. The method yields a convex relaxation that is numerically feasible and retains some state transition information. It is not guaranteed to produce the actual lower bound, it does allow for better approximations through higher order relaxations. The resulting convex problem is solvable using standard interior point methods.

The method developed herein is based on the following lemma from [2].

Lemma 1. *For a given cost function $J : \mathbb{R}^n \rightarrow \mathbb{R}$, the following two optimization problems are equivalent:*

$$\underset{x \in \mathcal{X}}{\text{minimize}} J(x) \iff \underset{\mu \in \Delta(\mathcal{X})}{\text{minimize}} \int_{\mathcal{X}} J(x) d\mu(x).$$

To illustrate how Lemma 1 is utilized, consider the following nonlinear, scalar difference equation for either candidate model:

$$x(k+1) = x(k) + x(k)^2 w(k). \quad (3)$$

Let the real state and disturbance variables $x(k)$ and $w(k)$ be the random variables $X(k)$ and $W(k)$ (which is without loss of generality in MDP and MIP by Lemma 1). Equation 3 can then be looked at as a moment relation, and, instead of propagating state, we can propagate state moments. This relaxes the constraint imposed by the difference equation and results in a lower bound on the cost in MDP. Higher moments can be computed to improve the bound but at the expense of computational costs.

To see how we can propagate the first moments of the state, let the reachable set of $x(k)$ lie in the set $\mathcal{X}(k) = [x^-, x^+]$ and let the domain of $w(k)$ be given by the set $\mathcal{W} = [w^-, w^+]$. Then take the expected value of both sides of the random difference equation to yield

$$\mathbb{E}[X(k+1)] = \mathbb{E}[X(k)] + \mathbb{E}[W(k)X(k)^2]. \quad (4)$$

At time k , the value of $\mathbb{E}[X(k)]$ is known since it was computed at time $k-1$. The value of $\mathbb{E}[W(k)X(k)^2]$ is not known, since we don't know the full distribution of $X(k)$. We can, however, constrain the value (using the theory of moments) since we know the reachable set $\mathcal{X}(k)$ and the disturbance domain \mathcal{W} .

The constraints on the mixed moment are given in [2] and briefly reviewed here in the following lemma. Some additional notation is required. For a random vector X and a positive integer N , let $X^{1 \times N}$ be the vector of all monomials up to the order N , e.g., $(X_1, X_2)^{1 \times 2} = (1, X_1, X_2, X_1^2, X_1 X_2, X_2^2)$, and let $X^{N \times N} = X^{N \times 1} X^{1 \times N}$. Furthermore, for an interval $[x^-, x^+]$ that contains the support of X , define the block diagonal matrix as $X_{\pm}^{N \times N} = \text{diag}((X_i - x_i^-)(x_i^+ - X_i) X^{N \times N})$.

Lemma 2. *The expected values of $X^{N \times N}$ and $X_{\pm}^{N \times N}$ are positive semidefinite.*

Lemma 2 is used to constrain the mixed moment in Equation 4 as follows. Denote the first moment of $\mathbb{E}[X]$ by \bar{x} . Expand the dimension of the disturbance space by assigning a new disturbance variable $\bar{w}_{\alpha_1, \alpha_2}$ to every monomial $\mathbb{E}[X^{\alpha_1} W^{\alpha_2}]$. Note that the disturbances corresponding to the first moment of X are not free, i.e., $\bar{w}_{1,0} = \bar{x}$. Finally form the matrices in Lemma 2 out of the disturbances \bar{w} and the state \bar{x} . What results is a convex approximation of Equation 3 with the form

$$\begin{aligned} \bar{x}(k+1) &= \bar{w}_{2,1}(k) + \bar{x}(k) \text{ (moment diff. eq.)}, \\ \bar{x}(k+1) &\in \mathcal{X}(k+1) \text{ (reachable stte)}, \\ V_N^{1,2}(k) &\geq 0 \text{ (moment matrix)}, \\ V_{N-1, \pm}^{1,2}(k) &\geq 0 \text{ (constraint matrix)}, \end{aligned} \quad (5)$$

where

$$\begin{aligned} V_N^{1,2}(k) &:= \mathbb{E} \left[(X(k), W(k))^{N \times N} \right], \\ V_{N-1, \pm}^{1,2}(k) &:= \mathbb{E} \left[(X(k), W(k))_{\pm}^{N \times N} \right]. \end{aligned}$$

To see how these constraints affect the transition of \bar{x} , see Figure 2. The figure plots a sample trajectory generated

from random disturbance values. At each point along the trajectory, we computed the set of possible values for the next state. These values are shown for the time $k = 7$. The moment constrained dynamics closely approximate the actual dynamics, while reachable set compute by solving the difference inclusion greatly overapproximates the actual dynamics.

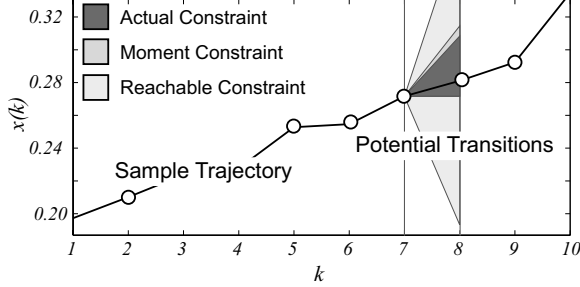


Fig. 2. A plot of a random trajectory for the simple model in Equation 3. At each point along the trajectory a set of possible future states can be computed using the moment constraint, the reachable set constraint, and the actual difference equation constraint. At time 7 the sets are shown in the figure.

The approach used for the system in Equation 3 can be directly extended for the compose model M , except now separate moment and constraint matrices will be formed for each monomial. Moreover, the matrices will have to also be conditioned on the control input trajectory. Recall that the transition function for a fixed control input u has the form $f(x, w) = \sum_{\alpha} a_{\alpha}(u) (x, w)^{\alpha}$. Allow the state and the disturbance to be random variables X and W , respectively, and denote the mixed moment $\mathbb{E}[(X, W)^{\alpha}]$ by \bar{w}_{α} . The domains of X and W are assumed to be some known intervals. For $i \in \{1, \dots, n\}$, let $\bar{w}_i = \bar{x}_i$, and let $A_u = (a_i(u))_{i \in \{1, \dots, n\}}$. Then, order the remaining disturbances (those not corresponding to first moments of X) into a vector \bar{w} and let B_u be the matrix with columns formed from the corresponding coefficients $a_{\alpha}(u)$. Then the difference equation conditioned on u^k can be approximated by the following convex constraints (denoted by $C(k; u^k)$)

$$C(k; u^k) \begin{cases} \bar{x}(k+1; u^k) = A_{u(k)} \bar{x}(k; u^{k-1}) \\ \quad + B_{u(k)} \bar{w}(k; u^{k-1}), \\ \bar{x}(k+1; u^k) \in \mathcal{X}(k+1; u^k), \\ \bar{y}(k; u^{k-1}) = C \bar{x}(k; u^{k-1}), \\ V_N^{\alpha}(k; u^{k-1}) \geq 0, \\ V_{N-1, \pm}^{\alpha}(k; u^{k-1}) \geq 0. \end{cases} \quad (6)$$

MDP with the exact difference equation constraint replaced by the constraints $C(k; u^k)$ is convex with respect to the disturbance. Moreover, if we fix the control input, the minimization problem is a linear matrix inequality problem, which can be solved efficiently. Finally, if we replace the exact difference equation constraint with the constraints $C(k; u^k)$ and then solve MDP the resulting minimization problem will yield a lower bound on the optimal cost. Therefore, if the optimal cost for the relaxed problem is positive for some U , then there exists a disparity certificate for the pair of models.

B. Disparity Certificate Derivation

The minimization problem with the relaxed transition constraints $C(k; u^k)$ can be solved using standard interior point methods. In this section, we address the maximization over the control inputs for the relaxed problem. It is important that we find the optimal control input since the relaxed minimization problem is already conservative. First we show how the minimization problem can be modified to simultaneously account for the maximization over the control inputs. Second we show how to iteratively compute the optimal control input.

Let MDP be the problem obtained by substituting the constraint $C(\cdot)$ for the difference equation constraint in MDP. This problem is convex in the disturbance variables when the probabilities are held fixed. Moreover, the cost function is concave (linear) with respect to the probabilities $P(u^K)$. Since the disturbances and probabilities belong to convex, compact sets, it is known that the problem has a saddle point, i.e., the max and min operations can be reversed [17]. Consider the following problem.

Convex Model Discrimination Problem (CoMDP). Given a composite model M , the problem is

$$\underset{\bar{w}, \lambda}{\text{minimize}} \lambda(1) + \|\bar{y}(1)\|_2^2,$$

subject to

$$\lambda(k; u^{k-1}) - \lambda(k+1; u^k) - \|\bar{y}(k+1; u^k)\|_2^2 \geq 0$$

$$C(k; u^k), \text{ for all } k \in \{1, \dots, K\}, u^k \in \mathcal{T}_K.$$

CoMDP is equivalent to the minimax problem where the control inputs can depend on the past and present values of the states and disturbances. The existence of a saddle point for MDP implies the following result.

Proposition IV.1. *The optimal cost of MDP' is equal to the optimal cost of CoMDP.*

CoMDP can be solved using standard interior point methods by forming a log barrier function for each inequality constraint. The resulting cost function at each centering step then has the form

$$J(\bar{w}) = \sum_{k=1}^{K+1} \sum_{u^{k-1} \in \mathcal{T}_{k-1}} \ell(\bar{x}(k; u^{k-1}), \bar{w}(k; u^{k-1})),$$

and is subject to the equality constraint $C(\cdot)$. Efficient stepwise Newton's method algorithms can then be used to minimize J . Such algorithms have the favorable property of linear complexity growth with respect to the number of time steps.

Next, we compute a control input U sufficient for discrimination. Let $P(u_k | u^{k-1})$ be the conditional probability that $U_k = u_k$, given that $U^{k-1} = u^{k-1}$. Let \bar{w}^* be the optimal disturbance value. Define the partial cost evaluated along the trajectory resulting from \bar{w}^* as

$$v_k(u^{k-1}) = \|\bar{y}(k; u^{k-1})\|_2^2 + \max_{u \in \mathcal{S}(u^{k-1})} v_{k+1}(u^{k-1}, u).$$

Define the set of maximizing control inputs as

$$\mathcal{Z}_k(u^{k-1}) = \underset{u}{\operatorname{argmax}} \{v_k(u^{k-1}, u)\}.$$

Then, a control input that will discriminate the pair of models is given in the next theorem.

Theorem 1. *Suppose the existence of a disparity certificate is shown by solving CoMDP, then the following control input will discriminate the pair of models with positive probability:*

$$P(u|u^{k-1}) = \begin{cases} \frac{1}{|\mathcal{Z}_k(u^{k-1})|}, & u \in \mathcal{Z}_k(u^{k-1}) \\ 0, & \text{otherwise} \end{cases}.$$

Proof: All we need to show is that the optimal control input must have the same support as U^K given by the transition probabilities $P(u|u^{k-1})$ described in the theorem. Suppose that for some k and u^{k-1} , the optimal control input assigns a positive probability to a u not in $\mathcal{Z}_k(u^{k-1})$. Then the expected cost will be less than v_1 , which contradicts Proposition IV.1 that says the optimal cost is equal to v_1 . ■ It is important to note that although Theorem 1 yields a control input that will discriminate the models, it is not the optimal control input. To compute the maximizing control input in MDP, further development is required.

Remark 1. If CoMDP produces a disparity certificate and different control inputs from the sample space of U in Theorem 1 are applied to the experimental system in multiple trials, then at least one model must be invalidated in a fine number of trial by solving MIP (MIP with the convex constraints $C(k; u^{k-1})$).

V. EXAMPLES

In this section we introduce two examples that illustrate the methods developed herein. A interior point method based on log barrier functions was developed that implements the stepwise Newton's method [18] to solve the centering problems. No special precautions were taken to deal with ill posed Hessian matrices, hence preventing large problems from being solved. It is important to note, however, that this is not a limitation of the method, rather a limitation of the current numerical implementation.

A. Example: Inhibition

Figure 3(a) shows a biochemical reaction network in with two species A and B . The input signal is the rate at which B is added and the output is the concentration of A . Two models of this network are presented in the table of Figure 3(a), the only difference being the range for the rate k_2 . The chemical kinetics model corresponding to the network is defined by

$$\begin{aligned} \dot{[A]} &= k_1 - k_2 [A][B] - k_3 [A], \\ \dot{[B]} &= u - k_2 [A][B]. \end{aligned} \quad (7)$$

We model the uncertainty in the reaction rates as a disturbance as in Equation 1, which implies that the reaction rates are allowed to vary with time. For the input u , we consider

impulses at specific time instances. To apply our method, we use the discrete time approximation of Equation 7 is

$$\begin{aligned} [A](k+1) &= [A](k) + \delta \left(k_1(k) \right. \\ &\quad \left. - k_2(k) [A](k) ([B](k) + u) - k_3(k) [A](k) \right), \\ [B](k+1) &= [B](k) + u(k) \\ &\quad + \delta \left(-k_2(k) [A](k) ([B](k) + u) \right). \end{aligned} \quad (8)$$

The maximum input is $u = 1$ and the minimum input is $u = .1$. Our method shows that the two models can be discriminated. The resulting disparity certificate is shown in Figure 3(b). For this system, a uniform distribute over input signals is enough to distinguish between the compete models. In Figure 3(c) we illustrate the output of model one for a random pair of inputs and show that in the worst case, model two cannot match the outputs (and hence is invalidated with respect to model one).

B. Example: Self-Annihilation

Figure 4(a) shows a network in which a species A activates its own degradation. The two candidate models for the system differ in the range they assign to the parameter k_2 , and thus to the degree to which A annihilates itself. Figures 4(b) and (c) are similar to those of the previous example, except that here the disparity certificate weights some input signals more than others, therefore suggesting a different array of experiments with the physical system.

VI. CONCLUSION

Model discrimination and model invalidation for polynomial system was considered in the framework of dynamic games. The class of disparate competing models was broadened through the use of stochastic inputs, i.e., models that no one input can discriminate, were shown to have a disparity certificate. The model invalidation problem for this setup was formulated. It was shown that at least one of two disparate models can be invalidated in a finite number of experiments. To solve the model invalidation and discrimination problems, a scalable convex relaxation method (based on the theory of moments) was developed and implemented in MATLAB. Future work includes a more efficient implementation of the developed theory and experimental application.

REFERENCES

- [1] J. T. Mettetal, D. Muzzey, C. Gmez-Urbe, and A. van Oudenaarden, "The frequency dependence of osmo-adaptation in *saccharomyces cerevisiae*," *Science*, vol. 319, no. 5862, 2008.
- [2] J. B. Lasserre, "Global optimization with polynomials and the problem of moments," *SIAM Journal on Optim.*, vol. 11, no. 3, 2001.
- [3] R. Horn, "Statistical methods for model discrimination. Applications to gating kinetics and permeation of the acetylcholine receptor channel," *Biophysical Journal*, vol. 51, no. 2, 1987.
- [4] R. Feely, M. Frenklach, M. Onsum, T. Russi, A. Arkin, and A. Packard, "Model discrimination using data collaboration," *Journal of Physical Chemistry A*, vol. 110, 2006.
- [5] J. F. Appgar, J. E. Toettcher, D. Endy, F. M. White, and B. Tidor, "Stimulus design for model selection and validation in cell signaling," *PLoS Comp. Biol.*, vol. 4, no. 2, 2008.

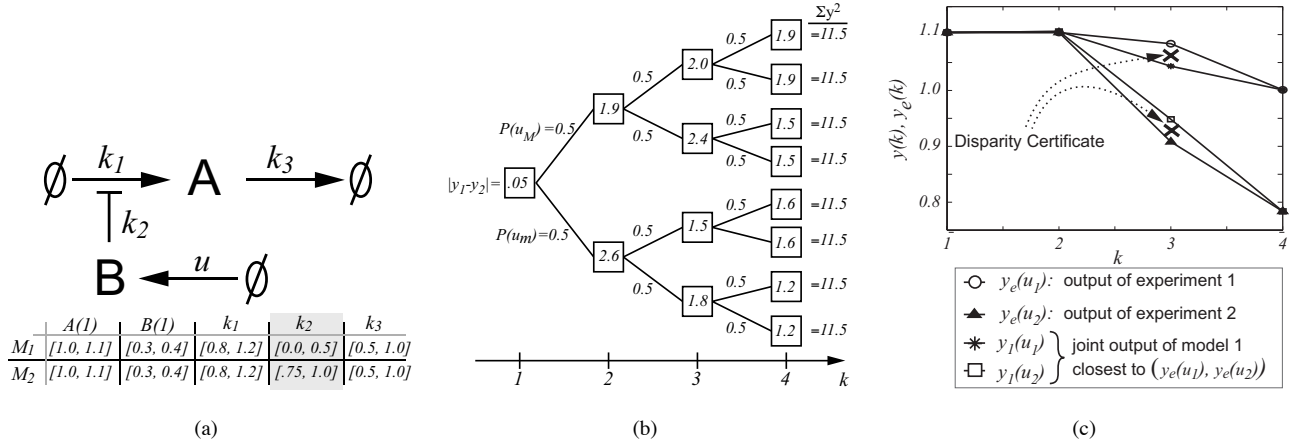


Fig. 3. (a) A simple biochemical network in which the production of a species A is inhibited by a species B . The input to the system is the rate at which B is added to the system. The two models for this system have identical structures, but different ranges for the rate k_2 (b) The computed disparity certificate. The tree traces all input trajectories. Each branch corresponds to a control input selection. Upper branches correspond to maximum input and lower branches correspond to minimum input. The transition probabilities corresponding to the disparity certificate are shown at their respective branches. For each time instance, the absolute value of the scaled (factor of 100) difference between y_1 and y_2 is specified. Following the leaves of the tree are shown the summed square costs. (c) Sample trajectories are of model one and the closest model two trajectory are shown. The plot indicates that if the physical system behaves according to model one, then model two will always be invalidated given enough experimental data.

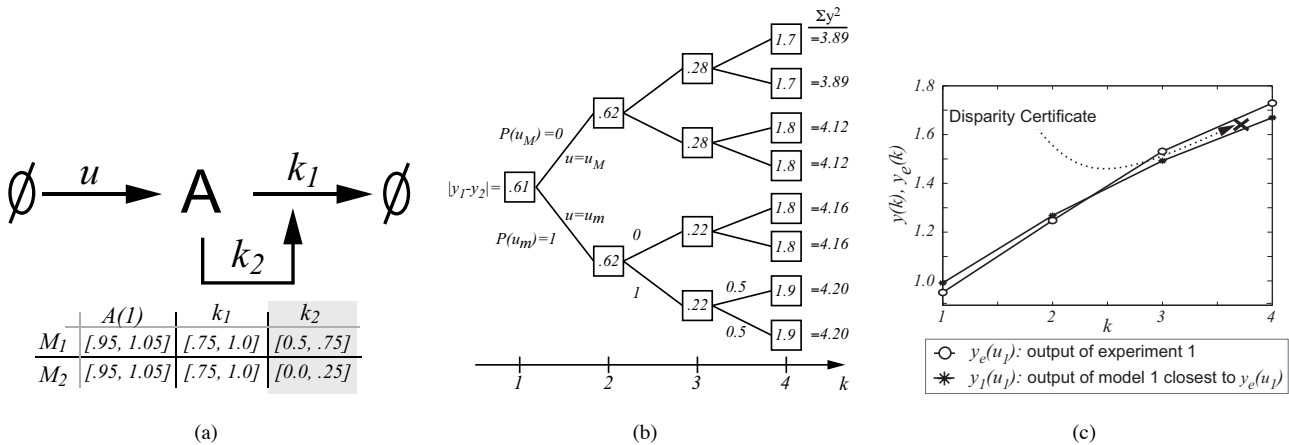


Fig. 4. (a) A biochemical network in which the degradation of a species A is activated by A itself. The two models for this system have identical structures, but different ranges for the rate k_2 (b) The computed disparity certificate. (c) Model one sample trajectories. The closest model two trajectory to the model one trajectory is also shown. The plot indicates that if the physical system behaves according to model one, then model two will always be invalidated given enough experimental data.

[6] A. Papachristodoulou and H. El-Samad, "Algorithms for discriminating between biochemical reaction network models: Towards systematic experimental design," in *Proc. of the American Control Conf.*, 2007.

[7] A. Kremling, S. Fischer, K. Gadkar, F. J. Doyle, T. Sauter, E. Bullinger, F. Allgower, and E. D. Gilles, "A benchmark for methods in reverse engineering and model discrimination: Problem formulation and solutions," *Genome Research*, vol. 14, 2004.

[8] L. Blackmore and B. Williams, "Finite horizon control design for optimal model discrimination," in *Proc. IEEE Conf. Decision & Control*, 2005.

[9] K. Felsenstein, "Optimal Bayesian design for discrimination among rival models," *Comp. Stat. and Data Anal.*, vol. 14, no. 4, 1992.

[10] T. Hatanaka and K. Uosaki, "Optimal auxiliary input for fault detection of systems with model uncertainty," in *Int. Conf. on Contr. App.*, 1999.

[11] S. Prajna, "Barrier certificates for nonlinear model validation," *Automatica*, vol. 42, no. 1, 2006.

[12] R. D. Nowak and B. D. V. Veen, "Random and pseudorandom excitation sequences for nonlinear system identification," in *IEEE Winter Workshop on Nonlinear Digital Signal Proc.*, 1993.

[13] R. B. Myerson, *Game Theory, Analysis of Conflict*. Harvard University Press, 1991.

[14] I. M. Mitchell, *Application of level set methods to control and reachability problems in continuous and hybrid systems*. PhD thesis, Stanford University, Dept. of Scientific Computing and Computational Mathematics, 2002.

[15] T. A. Henzinger, B. Horowitz, R. Majumdar, and W. Howard, *Lecture Notes in Computer Science*, ch. Beyond HYTECH: Hybrid Systems Analysis Using Interval Numerical Methods, pp. 130–144. Springer-Verlag, 2000.

[16] P. A. Parrilo, "Semidefinite programming relaxations for semialgebraic problems," *Mathematical Programming*, vol. 96, no. 2, 2003.

[17] T. Basar and G. Olsder, *Dynamic noncooperative game theory*. SIAM Classics in Applied Mathematics, second ed., 1999.

[18] J. C. Dunn and D. P. Bertsekas, "Efficient dynamic programming implementations of Newton's method for unconstrained optimal control problems," *Journal of Opt. Theory and App.*, vol. 63, no. 1, 1989.