Trajectory-based model reduction of nonlinear biochemical networks employing the observability normal form

Martin Löhnung, Jan Hasenauer, and Frank Allgöwer

Institute for Systems Theory and Automatic Control, University of Stuttgart, Pfaffenwaldring 9, Stuttgart, Germany, (e-mail: {loehnung,hasenauer,allgower}@ist.uni-stuttgart.de).

Abstract: In the last years, improved measurement devices, automated data generation, and high-throughput methods have initiated a trend to increasingly precise, but also increasingly complex models. Simulating these high-dimensional models and understanding their basic dynamic properties are crucial challenges encountered right now. One way towards these goals is model reduction. In this paper, we propose a trajectory-based method for reducing the input-output (I/O) map of continuous-time nonlinear ordinary differential equations. The method uses a sample of simulated I/O-trajectories, obtained by distributing initial states and input trajectories according to a probability density. Employing Monte-Carlo integration and the observability normal form, parameters of a reduced model are determined by convex optimization from this sample of trajectories. The properties of the method are illustrated using a model of the MAPK cascade. It is shown that redundancies are detected and that the approach can deal with nonlinear dynamics, such as limit-cycle oscillations.

Keywords: Model reduction, nonlinear models, observability normal form, convex optimization, MAPK cascade.

1. INTRODUCTION

The ultimate goals of systems biology are the prediction of treatment outcomes, the model-based development of targeted therapies, and the optimization of treatments on an individual basis (Anderson and Quaranta, 2008). To allow for this one has to be able to simulate the response to drugs on multiple spatial and temporal scales (Anderson and Quaranta, 2008; Parker and Clermont, 2010). Starting from the distribution of the drug within the whole body, taking place in minutes, down to the response of single cells, that may take days or weeks.

To achieve these aims simulation models for the different processes have to be developed that are predictive and easy to simulate. In particular at the single-cell scale a trend towards more and more detailed models (Klipp et al., 2005) can be observed. Unfortunately, these models are in general high-dimensional and highly nonlinear. Due to the high computational complexity, they do not allow for the integration into multi-scale models. On the other hand, these models provide highly quantitative predictions of the dynamic properties are crucial challenges encountered right now. One way towards these goals is model reduction. In this paper, we propose a trajectory-based method for reducing the input-output (I/O) map of continuous-time nonlinear ordinary differential equations. The method uses a sample of simulated I/O-trajectories, obtained by distributing initial states and input trajectories according to a probability density. Employing Monte-Carlo integration and the observability normal form, parameters of a reduced model are determined by convex optimization from this sample of trajectories. The properties of the method are illustrated using a model of the MAPK cascade. It is shown that redundancies are detected and that the approach can deal with nonlinear dynamics, such as limit-cycle oscillations.

In contrast, the basic idea behind trajectory-based model reduction is to (i) consider simulated I/O-trajectories and (ii) use system identification to determine a reduced model. Therefore, trajectory-based model reduction approaches do neither consider the states of the system nor operate on the nonlinear vector field $f$ of the ordinary differential equation (ODE) $\dot{x} = f(x, u)$. This avoids the analysis of the high-dimensional, nonlinear operators. Vargas and Allgöwer (2004) presented a procedure, using (i) and (ii), that is applicable to systems admitting a discrete-time Volterra representation, and suggested an iterative approach for the construction of the reduced model.

In this paper, we extend the work of Vargas and Allgöwer (2004) to reduced order continuous-time systems with polynomial right hand side. Additionally, weighting functions for initial conditions and input trajectories are introduced to allow for a design of tailored reduced models.

The approach we propose uses a weighted $\ell_2$-norm of the output error between reduced and original model as
objective function. As this objective function is high-dimensional it is evaluated using Monte-Carlo integration. To ensure a compact representation of the reduced model, it is assumed to be globally observable. This allows using the nonlinear observability normal form as representation. For the system identification of the reduced model a convex formulation is introduced. This convex formulation is inspired by the work of Niethammer et al. (2001), in which the particular structure of the observability normal form is employed. Finally, the complexity of the reduced model is decreased by enforcing sparsity.

The paper is structured as follows: In Section 2 we provide a detailed description of the problem setup. Following, the model reduction procedure is presented in Section 3. Section 4 contains a detailed case study of the MAPK cascade. The paper is concluded in Section 5.

2. PROBLEM STATEMENT

In this work, we consider the problem of model reduction for systems of nonlinear ODEs. The systems to be reduced will in the following be called detailed model and abbreviated with the subscript \( D \). For notational simplicity the detailed models are assumed to be a SISO system,

\[
\Sigma_D: \begin{cases}
    \dot{x} = f_D(x,u),
    x(0) = x_0 \\
y_D = h_D(x,u),
\end{cases}
\]

in which \( x(t,x_0,u) \in \mathbb{R}^n \) is the state of the system, \( u(t) \in \mathbb{R} \) is the input, and \( y_D(t,x_0,u) \in \mathbb{R} \) is the output. The initial condition is denoted by \( x_0 \in \mathbb{R}^n \). To ensure existence and uniqueness of solutions the vector field \( f_D : \mathbb{R}^n \times \mathbb{R} \to \mathbb{R}^n \) is assumed to be globally Lipschitz continuous. Furthermore, the mappings \( f_D \) and \( h_D : \mathbb{R}^n \times \mathbb{R} \to \mathbb{R} \) are considered to be sufficiently smooth. This is the case for most models of biological networks.

Given a detailed model \( \Sigma_D \), we are interested in a reduced model \( \Sigma_R \) that provides a good approximation of the I/O-behavior of \( \Sigma_D \) on a finite time interval \([0,T_{\text{end}}]\) and for a subset of initial conditions \( x_0 \) and input trajectories \( u \). Moreover, in many cases of practical relevance, certain initial conditions and/or input trajectories are known to be more likely or more important than others. Thus, we want to allow for the consideration of this prior knowledge during the model reduction. Such prior knowledge will in the following be encoded in importance weights. There are two independent weighting functions, \( p_x \) assigning importance to the initial condition and \( p_\varphi \) assigning importance to the input trajectory.

For ease of presentation, the input trajectories \( u_\varphi \in L_\infty \) are assumed to be parameterized by \( \varphi \in \mathbb{R}^{n_\varphi} \), e.g. a finite Fourier series, so it is possible to express the weighting as a function of the parameters \( \varphi \). \( L_\infty \) is the Lebesgue space of scalar-valued functions that are bounded on \([0,T_{\text{end}}]\).

For the weighting functions \( p_x : \mathbb{R}^n \to \mathbb{R}_+ \) and \( p_\varphi : \mathbb{R}^{n_\varphi} \to \mathbb{R}_+ \), we assume, that the integrals \( \int p_x(x_0)dx_0 \) and \( \int p_\varphi(\varphi)d\varphi \) exist and are equal to 1. Based on this, \( p_x(x_0) \) and \( p_\varphi(\varphi) \) can be interpreted as probability densities, where \( p_x(x_0) \) represents the probability of obtaining \( x_0 \) as initial condition and \( p_\varphi(\varphi) \) the probability of obtaining \( u_\varphi \) as input trajectory.

Remark: The weighting functions are design variables that allow for a tailored reduced model.

Based on this setup, the considered problem is:

Problem 1. Given a detailed model \( \Sigma_D \) and the relevant I/O-behavior defined by the weightings \( p_x(x_0) \) and \( p_\varphi(\varphi) \), compute a reduced model

\[
\Sigma_R: \begin{cases}
    \dot{x}_R = f_R(x_R,u_R),
    x_R(0) = g_R(x_0) \\
y_R = h_R(x_R,u_R),
\end{cases}
\]

of order \( n_R (x_R(t,x_0,u_\varphi) \in \mathbb{R}^{n_R}) \), such that the objective functional \( J(\Sigma_R) \) is minimized, with

\[
J(\Sigma_R) := \int_0^{T_{\text{end}}} E(x_0,u_\varphi) p_x(x_0) p_\varphi(\varphi) dx_0 d\varphi,
\]

in which \( E \) is the integrated square output error

\[
E(x_0,u_\varphi) := \int_0^{T_{\text{end}}} (y_D(t,x_0,u_\varphi) - y_R(t,x_0,u_\varphi))^2 dt.
\]

In short, the goal is to find, for a given order \( n_R \) of the reduced model, the vector field \( f_R \), the output mapping \( h_R \) and the initial state mapping \( g_R \), such that weighted integrated squared output error is minimal.

3. TRAJECTORY-BASED MODEL REDUCTION

In this section we show a novel procedure to address Problem 1 by determining a reduced model employing Monte-Carlo integration, the observability normal form, and convex optimization.

3.1 Evaluation of the objective function

The basic idea of trajectory-based model reduction is the usage of simulated trajectories for the assessment of the reduced model. Accordingly, the objective functional \( J(\Sigma_R) \) is determined using simulation. As the integral defining \( J(\Sigma_R) \) is high-dimensional \((1 + n + n_\varphi)\) classical integration methods are not applicable.

In this work we employ Monte-Carlo integration (MacKay, 2005) to overcome this problem. To determine an approximation of \( J(\Sigma_R) \), a sample of initial conditions \( \{x_0^{(i)}\}_{i=1}^S \) and input trajectories \( \{u_\varphi^{(i)}\}_{i=1}^S \) is drawn, with \( x_0^{(i)} \sim p_x(x_0) \) and \( u_\varphi^{(i)} := u_\varphi(x_0) \) with \( \varphi^{(i)} \sim p_\varphi(\varphi) \). The notation \( z^{(i)} \sim p_z(z) \) denotes that the sample member \( z^{(i)} \) is drawn according to \( p_z(z) \). The number of sample members is denoted by \( S \). Furthermore, a sample of time points \( \{t^{(i)}\}_{i=1}^S \) is drawn from a uniform distribution over \([0,T_{\text{end}}]\). Given \( \{x_0^{(i)}\}_{i=1}^S, \{u_\varphi^{(i)}\}_{i=1}^S, \) and \( \{t^{(i)}\}_{i=1}^S \), a sample of points from I/O-trajectories is given by

\[
T := \left\{ y_D^{(i)}(t^{(i)}, x_0^{(i)}, u_\varphi^{(i)}) \right\}_{i=1}^S.
\]

From classical Monte-Carlo integration it is now known that for \( S \gg 1 \) a good approximation of \( J(\Sigma_R) \) is

\[
\hat{J}(\Sigma_R) := \sum_{i=1}^S \left( y_D^{(i)}(t^{(i)}, x_0^{(i)}, u_\varphi^{(i)}) - y_R^{(i)}(t^{(i)}, x_0^{(i)}, u_\varphi^{(i)}) \right)^2.
\]

Given the approximated, evaluable objective function \( \hat{J}(\Sigma_R) \), we are left with the problem of minimizing \( \hat{J}(\Sigma_R) \) over \( f_R, h_R, \) and \( g_R \).
3.2 Parameterization of the reduced model

In order to keep the optimization of the reduced model trackable, the model and input class is restricted slightly. We assume that the reduced model $\Sigma_R$ is (i) globally observable. Additionally, (ii) only $n_R$ times continuous differentiable inputs $u_\varphi$ are allowed.

Given that (i) and (ii) holds, each possible reduced model can be represented using its observability normal form (Zeitz, 1984),

$$\dot{\xi}_j = \dot{\xi}_{j+1}, \quad j = 1, \ldots, n_R - 1$$
$$\dot{\xi}_{n_R} = \nu(\xi, \mu)$$
$$y_R = \xi_1$$

(2)

with the extended input vector $\mu := [u_\varphi, \dot{u}_\varphi, \ldots, (n_R)T]$ and function $\nu : \mathbb{R}^{n_R} \times \mathbb{R}^{n_R+1} \rightarrow \mathbb{R}$. In this unique state representation, the states of the reduced model are the time derivatives of the output, $\xi = [y_R, \dot{y}_R, \ldots, (n_R)^T]$. This is beneficial as also a direct link between the states $\xi$ and the output of the detailed model is established, which e.g. makes a reasonable choice for $g_R$.

$$\xi_k(0) := (k-1) y_D(0, x_0, u_\varphi), \quad k = 1, \ldots, n_R.$$  

Note that assumptions (i) and (ii) are crucial but not restrictive. In particular, (i) ensures that the reduced model has no hidden states. The more critical assumption is (ii), but for biological systems inputs are in general smooth, as most processes are diffusion driven.

Representation (2) facilitates the identification of a reduced model. Instead of identifying the vector-valued functions $f_R$ and $g_R$ and the scalar-valued function $h_R$, we are left with the problem of identifying a single scalar-valued function $\nu$. Nevertheless, identifying $\nu$ still requires an optimization over a function space.

To obtain a parameterized model, $\nu(\xi, \mu)$ is restricted to be a polynomial with degree $d$,  

$$\nu_d(\xi, \mu, \theta) := \theta^T m_d(\xi, \mu),$$

where $\theta \in \mathbb{R}^n$ is the vector of coefficients and $m_d(\xi, \mu)$ denotes the vector of all monomials with degree less or equal $d$, i.e.

$$m_d(\xi, \mu) = [1, \xi_1, \ldots, \mu_{n_R+1}, \xi_1^2, \xi_2, \ldots, \mu_{n_R+1}^d]^T.$$  

The number of monomials is $n_\theta = \binom{n + d}{d}$.

Given the parameterized function $\nu_d$, the problem of estimating $\Sigma_R$ is reduced to determining an appropriate order $d$ and estimating $\theta$. Hence, for fixed $d$, $\hat{J}(\Sigma_R)$ is only a function of $\theta$, $\hat{J}(\theta)$.

3.3 Estimation of the reduced model

Classical nonlinear optimization: One method to estimate the parameters $\theta$ is classical nonlinear optimization. This approach is the predominate approach for nonlinear continuous-time parameter estimation.

In this model reduction context, this classical approach requires the minimization problem

$$\min \sum_{i=1}^S \left( y_D^{(i)}(t^{(i)}, x_0^{(i)}, u_\varphi^{(i)}) - y_R^{(i)}(t^{(i)}, x_0^{(i)}, u_\varphi^{(i)}) \right)^2,$$

subject to $\xi_{j+1}^{(i)} = \xi_{j+1}^{(i)}$, $j = 1, \ldots, n_R - 1$,  

$$\xi_{n_R}^{(i)} = \nu_d(\xi_{n_R}^{(i)}, \mu^{(i)}),$$

$$\xi_k^{(i)}(0) := (k-1) y_D^{(i)}(0, x_0^{(i)}, u_\varphi^{(i)}), \quad k = 1, \ldots, n_R,$$

$$y_R^{(i)} = \xi_1^{(i)}, \quad i = 1, \ldots, S,$$  

(3)

to be solved. As this problem is in general nonlinear and non-convex, computation of the global optimum cannot be ensured. Additionally, for algorithms seeking the global optimum the computational effort grows in general tremendously with $S$.

Convex optimization approach: In this paper, another approach is proposed that exploits the fact that model reduction is performed and not classical parameter estimation from measurement data.

While performing model reduction, not only the output $y_D$ but also the derivatives of $y_D$ are available. Considering the reduced model, this means that the target values for all states $\xi$ of the reduced model are known. In particular we have a target value for $\xi_{n_R}$, the $n_R$-th derivative of the output $y_D$. If we could ensure that

$$\forall t, \forall x_0, \forall u_\varphi : \left( y_R(t, x_0, u_\varphi) = y_D(t, x_0, u_\varphi) \right),$$

holds, the I/O-mapping of the detailed and reduced model would match exactly. This inspires a modification of the objective function to

$$\hat{J}_\nu(\theta) := \sum_{i=1}^S \left( y_D^{(i)}(t^{(i)}, x_0^{(i)}, u_\varphi^{(i)}) - \nu_d(\xi_{n_R}^{(i)}, \mu^{(i)}, \theta) \right)^2,$$

with

$$\xi_{n_R}^{(i)} := \left[ y_D^{(i)}(t^{(i)}, x_0^{(i)}, u_\varphi^{(i)}), \ldots, (n_R-1)^{(i)}(t^{(i)}, x_0^{(i)}, u_\varphi^{(i)}) \right]^T$$

$$\mu^{(i)} := [u_\varphi^{(i)}, \dot{u}_\varphi^{(i)}, \ldots, (n_R)^{(i)}(t^{(i)}, x_0^{(i)}, u_\varphi^{(i)})]^T.$$  

This modified objective $\hat{J}_\nu(\theta)$ is closely related to $\hat{J}(\theta)$. The main difference is that $\hat{J}_\nu(\theta)$ does not penalize the integrated error along the trajectories. Instead, only the point-wise error in the evolution of the detailed and reduced model is penalized. This can cause a time shift between $y_D$ and $y_R$ if the long-term behavior is considered.

Given $\hat{J}_\nu(\theta)$ we obtain the modified optimization problem:

$$\min \hat{J}_\nu(\theta).$$  

(4)

Solving (4) is equivalent to computing the parameterization $\theta$ of $\nu_d$, that results in the smallest difference $\| y_D - y_R \|_2$, assuming that the reduced model is in the correct point in the space of outputs and derivatives thereof.

Compared to the classical nonlinear optimization problem (3) the formulation (4) has one big advantage. As $\xi_{D}^{(i)}$ and $\mu^{(i)}$ are determined in advance, we can rewrite (4) to

$$\text{minimize } \| A \theta - b \|_2^2,$$

with $A := [a_1, \ldots, a_S]^T$, $a_i := m_d(\xi_d^{(i)}, \mu^{(i)})$, and $b := [b_1, \ldots, b_S]^T$, $b_i := (n_R)^{(i)}(t^{(i)}, x_0^{(i)}, u_\varphi^{(i)})$. This clearly shows that (4) is a convex quadratic optimization problem.
3.4 Complexity reduction of reduced model

Although (4) allows for an efficient computation of $\theta$ one problem remains. In general, all elements of $\theta$ are non-zero, resulting in a reduced model with $n_\theta = 2^{n_M + 1 - \delta}$ monomials. Such a complex reduced model is often unnecessary. It is desirable to include only the required monomials by forcing $\theta$ to have many zero elements.

To enforce sparsity of $\theta$ a method known from compressive sensing (Candes et al., 2006; Donoho, 2006) can be employed. In this work a $\ell_1$-norm formulation is used that results in sparse solutions. The corresponding convex optimization problem is

$$\minimize \| \tilde{\theta} \|_1$$

subject to

$$\tilde{J}_\nu(W^{-1}\tilde{\theta}) \leq \lambda \tilde{J}_\nu(\theta^*)$$

in which $W$ is a diagonal matrix with the $\ell_2$-norm of the columns of $A$ on the diagonal, $\theta = W^{-1}\tilde{\theta}$, and $\theta^*$ is the solution of (4). The normalization with $W$ ensures that the elements of $\tilde{\theta}$ are comparable. The parameter $\lambda$ restricts the increase of the cost function and thus the decrease in approximation accuracy. Hence, the choice of $\lambda$ is critical and the trade off between model complexity and approximation quality has to be kept in mind.

To summarize, we presented an approach to derive reduced order models in observability normal form, by convexifying Problem 1. Employing $\ell_1$-minimization the complexity of the resulting models can be further reduced.

4. EXAMPLE: MAPK CASCADE

In this section some important properties of the proposed model reduction approach are illustrated by studying the mitogen-activated protein kinase (MAPK) cascade. The MAPK cascade has been chosen due to its importance, as it represents a frequently appearing motive in biological pathways, and as mainly the I/O-behavior is of interest, making it perfectly suited for our approach.

4.1 Introduction

The MAPK cascade is one of the most intensively studied pathway in eukaryotic cells (Orton et al., 2005; Schöberl et al., 2002). It plays an important role in many signaling processes and is crucially involved in different cell fate decisions such as growth, proliferation, differentiation, and cell survival.

The term 'MAPK pathway' does not refer to a single pathway but to commonly seen signaling modules in eukaryotic cells (Orton et al., 2005). These modules have very similar structures and in general consist of three kinases, e.g. Raf, MEK and ERK, activated by sequential phosphorylation. The whole signaling process is initiated by the binding of ligands, like the EGF (Schöberl et al., 2002), to cell surface receptors. This results in the activation of the downstream MAPK cascade, that amplifies the signal. Finally, the active form of the last kinase translocates into the nucleus, yielding a change in transcription factor activity. The subsequent change of protein concentration influences many different processes.

4.2 MAPK cascade model and its properties

For this study, the model proposed by Khodenko (2000) has been chosen because this model is well established. The whole model has eight state variables, three kinases in different phosphorylation states as illustrated in Figure 1.

The input to the MAPK cascade is the activity of the membrane-bound GTPase Ras, $u_\psi(t) = c \cdot [\text{Ras}](t)$, that catalyzes the phosphorylation of Raf. The factor $c$ is the normalization constant ensuring $u_\psi(t) \in [0, 1]$. The output is the concentration of double phosphorylated ERK $y_D(t) = [\text{ERK-PP}](t)$.

The differential equations governing the time evolution of the state variable can be found in (Khodenko, 2000). All reaction rates are modeled by first or second order Hill kinetics, hence rational functions. In (Khodenko, 2000) different parameters as well as different feedback mechanism have been investigated. Due to space constraints, we consider only the case of cooperative inhibition for which the parameter values can be found in (Khodenko, 2000, caption of Figure 2). The unit for $t$, $y$ is min, nM respectively, and is omitted in the following.

4.3 Model reduction

To perform the model reduction, the data is generated by simulating the detailed MAPK cascade model. The probability density over initial conditions $p_\psi(x_0)$ has been chosen to be a sum of two Gaussian distributions. The mean values of these Gaussian distributions are the centers of the limit cycles that are exhibited by the system for $u_\psi = 0.1$ and $u_\psi = 1$. The standard deviations of the initial conditions are 5% and 10% of the upper limit of the states.

Due to conservation relations the initial conditions are limited to a linear subspace with equal absolute amount of the three kinases, e.g. $\text{Raf} + \text{Raf-P} = \text{Raf}_{\text{tot}}$. The input trajectories are chosen to be step functions smoothed by the transfer function $G(s) = 1/((100s + 1)^{n_M})$,

$$u_\psi(t) = \mathcal{L}^{-1}(G(s)) * \tilde{u}_\psi(t) + \tilde{u}_1 \mathcal{L}^{-1} \left( \frac{(100s + 1)^{n_M} - 1}{s(100s + 1)^{n_M}} \right),$$

$$\tilde{u}_\psi(t) = \begin{cases} \tilde{u}_1, & t \leq \tau \\ \tilde{u}_2, & t > \tau \end{cases}$$

Fig. 1. Illustration of the MAPK pathway model proposed in (Khodenko, 2000), containing conversion reactions (---), regulatory interactions (\longrightarrow), inhibition (\rightarrow), and the different species.
with the convolution operator $*$, Laplace transform $\mathcal{L}$, and $\varphi = [\bar{u}_1, \bar{u}_2, \tau]^T$. The probability density over inputs is

$$ p_\varphi(\varphi) = \Phi(\bar{u}_1[0, 1])\Phi(\bar{u}_2[0, 1])\Phi(\tau[0, 150]), $$

in which $\Phi(x|\alpha, \beta)$ denotes the probability density of the uniform distribution in the interval $x \in [\alpha, \beta]$. For the model reduction a sample with 2000 members is drawn from $p_\varphi(x_0), p_\varphi(\varphi)$, and $p_\tau(t) = \Phi(t[0, 150])$ respectively. For the resulting triples the MAPK cascade is simulated, yielding the sample of I/O-trajectories $T$.

Given the sample $T$, the order $n_R$ of the reduced model has to be chosen. As the detailed model has a relative degree three, $n_R$ is set to three. Due to the nonlinear kinetics of the detailed model, a polynomial degree of five is chosen. This yields the reduced model with up to 792 parameters and monomials,

$$ \Sigma_R: \begin{cases} \dot{\xi}_1 &= \xi_2 \\ \dot{\xi}_2 &= \xi_3 \\ \dot{\xi}_3 &= \nu_3(\xi, \mu, \theta) = \theta^T m_3(\xi, \mu) \\ y_R &= \xi_1. \end{cases} $$

Based on this structure of the reduced model and the sample $T$, a parameterization of the reduced model is determined by solving (5), using the MATLAB toolbox Yalmip (Löfberg, 2009) with the solver SeDuMi (Sturm, 1999). By changing the optimization parameter $\lambda$, we can thereby reduce the number of numerical non-zero elements in $\theta$, as depicted in Figure 2. The relation is indicated by the sharp bend in the absolute value of $\tilde{\theta}_i$. A trade-off between model accuracy and complexity is achieved with $\lambda = 1.5$, resulting in 61 monomials. Note that the large number of required monomials is partially due to the rational functions contained in the original model.

### 4.4 Analysis and validation of the reduced model

To validate the reduced model, the approximation error of the mapping $e(\xi) = \tilde{y}_D(\xi) - \nu_3(\xi_D(\xi), \mu(\xi), \theta)$ is evaluated for a validation sample. Therefore, Figure 3 depicts the standard deviations $\sigma_e$ of $e(\xi)$ and $\sigma_\nu$ of $|\nu_3(\xi_D(\xi), \mu(\xi), \theta)|$ of the validation sample. It is apparent that the standard deviation of the approximation error $e$ decays rapidly. Furthermore, it is more than one order of magnitude smaller than the standard deviation of $\nu_3(\xi_D(\xi), \mu(\xi), \theta)$, when the model is not excited. The larger values after $t = 0$, respectively $t = 75$ are due to the initial conditions and/or input signals. The convergence of the trajectories to the nonlinear manifold defined by $y_D - \nu_3(\xi_D(\xi), \mu(\xi), \theta) = 0$ suggests that the long term behavior can be reproduced well.

The difference of the output trajectory between the detailed and reduced model is not directly related to $y_D - \nu_3(\xi_D(\xi), \mu(\xi), \theta)$. Thus, trajectories from simulations with $x_0 \sim p_x(x_0)$ are used as further validation. EXEMPLARY. trajectories for increasing $(\varphi = [0, 1, 1, 75])$ and decreasing $(\varphi = [0.9, 0.2, 75])$ steps as input signal are shown in Figure 4. It is obvious, that the quantitative behavior is reproduced well by the reduced model.

To emphasize the necessity of a nonlinear model reduction procedure, we also compute a linear model with three states using our procedure, i.e. $\nu_1(\xi, \mu, \theta) = \theta^T [\xi_1, \xi_2, \xi_3, u_\varphi, \bar{u}_2, \bar{u}_\varphi, \bar{u}_\varphi]$. This yields a large decrease in performance.

The existence of limit cycles is a major property of the MAPK cascade model, stressing its nonlinearity. To visualize the limit cycle, the trajectories of the outputs for $t \in [0, 1500]$ and $\varphi = [0, 0.9, 75]^T$, projected onto the $y-y_\varphi$-plane, are shown in Figure 5. The limit cycle is captured by the reduced model.

Besides the limit cycle, Figure 5 also shows the points used for estimation of the mapping $\nu_3(\xi, \mu, \theta)$ projected onto the $y-y_\varphi$-plane, i.e. $\{[\bar{y}_D(t(i), x_0(i), u_0(i)), \bar{y}_D(t(i), x_0(i), u_0(i))]\}_{i=1}^{58}$.\textsuperscript{8} An analysis of this set unravels that the trajectory of $y_D$ for $u_\varphi = 0$ is close to the border of supporting points. This makes the estimation of the mapping $\nu(\xi, \mu, \theta)$ difficult around the steady state for $u_\varphi = 0$. One result of this is, that the reduced model return negative values for $y_R$.\textsuperscript{9}
of freedom and the utilization of the problem structure to obtain a convex optimization problem.

The properties of the method are illustrated using a model of the MAPK cascade. We show that the model reduction procedure can handle nonlinear behavior like limit cycles. Furthermore, the proposed approach yields an order reduction and a reduced complexity model.

Future work is underway to test the proposed procedure for large-scale models. Due to the fact, that the key connection between the detailed and the reduced model are the output trajectories, advantages compared to projection-based model reduction procedures are expected.

REFERENCES


