Preprints of the 21st IFAC World Congress (Virtual)
Berlin, Germany, July 12-17, 2020

Data-driven surrogate models for LTI systems via saddle-point dynamics ⋆

Tim Martin, Anne Koch, Frank Allgöwer

Institute for Systems Theory and Automatic Control, University of Stuttgart

Abstract: For the analysis, simulation, and controller design of large-scale systems, a surrogate model with small complexity is mostly required. A standard approach to determine such a model is given by modelling the system and applying model-order-reduction techniques. Contrary, we propose a data-driven approach, where the surrogate model of the input-output behaviour of an LTI system is determined from data without modelling the system beforehand. Moreover, we provide a guaranteed bound on the maximal error between the system and the surrogate model in case of noise-free measurements. We analyse the stability and convergence of the presented schemes and apply them on a benchmark system from the model-order-reduction literature.

Keywords: iterative methods, reduced-order models, input-output methods, learning algorithms, gradient methods

1. INTRODUCTION

Deriving a low-order surrogate model has great importance in simulation, analysis, and controller design of large-scale systems, e.g., mechanical models of flexible multibody systems or weather-forecast models. Model-order-reduction techniques constitute one approach to calculate such models. In recent decades, different model-order-reduction techniques have been investigated as projection-based, norm-based (e.g. $\mathcal{L}_2, \mathcal{L}_\infty, \mathcal{H}_2, \mathcal{H}_\infty$), and moment matching approaches. See Antoulas (2005) for more details. In model order reduction, the knowledge of linear or nonlinear ordinary differential equations, that describe the system dynamics, is required. These equations are derived possibly by discretizing partial differential equations models using, for instance, a finite element method. Subsequent, the model reduction step yields a simplified dynamical system that approximates the behaviour of the high-order model. Here, a trade-off between complexity and misfit in the approximation has to be made.

As stated above, the starting point of model order reduction is a sufficiently precise model of the system. However, a model for large-scale systems is often not available and modelling is time-consuming. For these reasons, approaches for data-driven surrogate modelling were developed, e.g., in Ionita and Antoulas (2014) and Scarciotti and Astolfi (2017). Instead of modelling the system and applying model-order-reduction techniques, a simplified model is deduced directly from data of the system. Thereby, one goal of data-driven approaches is to provide a bound of the approximation error with respect to the real plant. In contrast, model order reduction can only provide such a bound regarding the high-order model, while the error of this high-order model is unknown.

⋆ This work was funded by Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under Germany's Excellence Strategy - EXC 2075 - 390740016. The authors thank the International Max Planck Research School for Intelligent Systems (IMPRS-IS) for supporting Anne Koch. For correspondence, tim.martin@ist.uni-stuttgart.de.

In this paper, we establish a data-driven approach for surrogate modelling of LTI SISO systems. We consider a similar approach as Wahlberg et al. (2010) and Romer et al. (2018), where optimization problems are solved iteratively to determine the ℓ_2-gain and conic relations of LTI systems without the knowledge of the system. Instead, the schemes require to perform sequentially experiments on the plant to identify the control-theoretic properties. In Oomen et al. (2014), such an approach is successfully applied on a real industrial active vibration isolation system in the case of $\mathcal{L}_2$-gain estimation. Besides a surrogate model, we compute the ℓ_2-gain of the approximation error between the system and the surrogate model. Thus, a measure of the misfit between the surrogate model and the real plant is provided under the assumption of noise-free measurements. Moreover, the bound on the maximal error can be exploited for a robust control design with closed-loop guarantees. Since the surrogate model can be seen as a generalisation of conic relations from Zames (1966) with dynamic center, this robust controller can exhibit a better performance than a controller by the classical feedback theorem from Zames (1966). Due to this generalisation, we retrieve the result from Romer et al. (2018) as a special case. Note that, set-membership identification from Milanese et al. (2014) constitutes an approach to derive a model of a nonlinear system and its maximal approximation error. However, these results are not applicable here, since the linear surrogate model constitutes the projection of the (high-order) linear system on a set of (low-order) linear models, which follows a different idea as in Milanese et al. (2014).

The paper is organized as follows. First, we state the characterizing optimization problem of the 'optimal' surrogate model. Second, we solve this problem by continuous-time saddle-point dynamics. Subsequent, the stability and convergence is analysed, which also provides insight into the convergence behaviour of the iterative schemes. At the end, we apply these iterative schemes on a benchmark system from the model-order-reduction literature.

2. SURROGATE MODELS FOR LTI SYSTEMS

In this paper, we consider causal discrete-time LTI SISO systems and assume that $u(t) = y(t) = 0$ for $t \leq 0$. Hence, the output of the system at each time step t can be computed by

$$y(t) = \sum_{k=0}^{t} g_k u(t-k)$$

and a finite output sequence by

$$\begin{bmatrix} y(1) \\ y(2) \\ y(3) \\ \vdots \\ y(n) \end{bmatrix} = \begin{bmatrix} g_0 & 0 & 0 & \cdots & 0 \\ g_1 & g_0 & 0 & \cdots & 0 \\ g_2 & g_1 & g_0 & \cdots & 0 \\ \vdots & \vdots & & \ddots & \vdots \\ g_{n-1} & g_{n-2} & \cdots & g_1 & g_0 \end{bmatrix} \begin{bmatrix} u(1) \\ u(2) \\ u(3) \\ \vdots \\ u(n) \end{bmatrix}.$$

In the sequel, we use the abbreviation $y = Gu$ with $y, u \in \mathbb{R}^n$ and the lower triangular Toeplitz matrix $G \in \mathbb{R}^{n \times n}$. The surrogate model is restricted to be a linear combination of finitely many LTI systems. Thus, its convolution matrix $T(b)$ is a linear combination of lower triangular Toeplitz matrices $\tilde{T}_i, i = 1, \ldots, N$

$$T(b) = \sum_{i=1}^{N} b_i \tilde{T}_i, \tag{1}$$

with $b = [b_1 \ \cdots \ b_N] \in \mathbb{R}^N$. Later, the linearity of $b \mapsto T(b)$ will be crucial to prove stability guarantees and to provide a fast converging scheme.

To motivate this setup, we consider the following conceivable problem. Suppose the system is an IIR filter of high order M

$$y(t) = \frac{\beta_0 + \beta_1 q^{-1} + \cdots + \beta_M q^{-M}}{1 + \alpha_1 q^{-1} + \cdots + \alpha_M q^{-M}} u(t), \alpha_M \neq 0,$$

with unknown parameters $\alpha_i, i = 1, \ldots, M$ and $\beta_i, i = 0, \ldots, M$. For the design of a low order controller, we want to find an approximating low-order IIR filter

$$y(t) = \frac{b_0 + b_1 q^{-1} + \cdots + b_N q^{-N}}{1 + a_1 q^{-1} + \cdots + a_N q^{-N}} u(t), a_N \neq 0,$$

with $N \ll M$. Here, the parameters $b_i, i = 0, \ldots, N$ are obtained through optimization, whereas the parameter $a_i, i = 1, \ldots, N$ are chosen by initial experiments and some insight into the system beforehand. Thereby, $a_i, i = 1, \ldots, N$ can be seen as design parameter of the iterative scheme.

The goal of this work is to determine the 'best fitting' surrogate model, which we define by the optimization problem

$$\Delta^2 := \min_{b \in \mathbb{R}^N} \max_{u \in \mathbb{R}^n \setminus \{0\}} \frac{||(G - T(b))u||^2}{||u||^2}. \tag{2}$$

$|| \cdot ||$ denotes the Euclidean vector norm and (b^*, u^*) the solution of (2). Through this optimization problem, the 'optimal' surrogate model $T(b^*)$ minimizes the gain Δ of the error system $G - T(b)$ with respect to the Euclidean vector norm. Therefore, the surrogate model $T(b^*)$ minimizes the distance of surrogate models $T(b), b \in \mathbb{R}^N$ to the system G and $T(b^*)$ can be seen as the orthogonal projection of G on the set of the surrogate models given by $T(b), b \in \mathbb{R}^N$. This is illustrated in Fig. 1.

In the following, we comment on the relation of (2) to other results from the control theory. First, by the knowledge of b^* and Δ, the unknown system G can be represented as interconnection as depicted in Fig. 2. Therefore, a robust controller design with closed-loop guarantees is conceivable. Second, the optimization problem (2) includes the definition of conic relations with static center from Zames (1966) for $T(b) = bI$ and the definition of the ℓ_2-gain of G for $b = 0$. Hence, the robust controller determined from the surrogate model $T(b^*)$ with error bound Δ can lead to a better performing controller than by identifying the minimal cone, that includes G, and applying the feedback theorem from Zames (1966). Indeed, the width of the cone containing G for a static center is larger than for a dynamic center. Thus, a stabilizing controller, concluded from a static center, is confined in a smaller cone, and hence is more restricted. Third, we can relate the optimization problem (2) to nonlinearity measures, where G is a nonlinear system. In this case, Δ measures the strength of the nonlinearity of a system. In Martin and Allgöwer (2019), nonlineaity measures are investigated for data-driven system analysis.

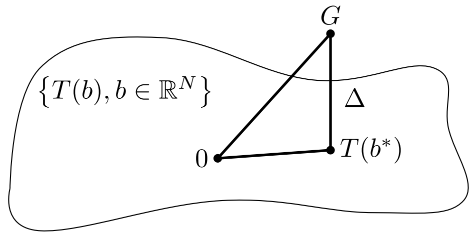

Fig. 1. 'Best fitting' surrogate model as the orthogonal projection of system G on the set $\{T(b), b \in \mathbb{R}^N\}$.

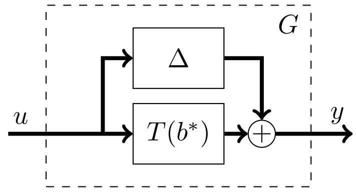

Fig. 2. Unknown system G as interconnection of the 'best fitting' surrogate model and the bound Δ on the error.

In the following, we require some more notation related to (2). We define

$$\frac{u^T(G - T(b))^T(G - T(b))u}{u^T u} =: \frac{u^T A(b) u}{u^T u} =: \rho(b, u).$$

ρ is known as the Rayleigh quotient and is a smooth function with $\rho(b, u) = \rho(b, \alpha u)$ for all $\alpha \in \mathbb{R} \setminus \{0\}$. Thus, the considered inputs in the optimization can be restricted to the unit sphere $S^{n-1} \subset \mathbb{R}^n$.

Since $A(b)$ is symmetric, it is known that

$$\max_{||u||=1} \frac{u^T A(b) u}{u^T u} = \lambda_{\max}(b) \tag{3}$$

and that the maximizing input $u^*(b)$ corresponds to any linear combination of eigenvectors of the largest eigenvalue of $A(b)$. Here, $\lambda_{\max}(b) = \max_{i=1,\ldots,n} \lambda_i(b)$ denotes the maximal eigenvalue of $A(b)$ and $\lambda_i(b), i = 1, \ldots, n$ its i-th eigenvalue. Hence, the optimization problem (2) can be written as

$$\Delta^2 = \min_{b \in \mathbb{R}^N} \rho(b, u^*(b)) = \min_{b \in \mathbb{R}^N} \lambda_{\max}(b). \tag{4}$$

By Mengi et al. (2014), we can sort the eigenvalues of $A(b)$ such that the eigenvalue functions $\lambda_i(b) : \mathbb{R}^N \to \mathbb{R}, i =$

$1,\dots,n$ are analytic in b. This implies the existence of b^* and the first order necessary conditions for the solution (b^*, u^*) of (2)

$$\mathrm{conv}\{\nabla_b\lambda_1(b^*),\dots,\nabla_b\lambda_m(b^*)\} \ni 0,$$

where conv denotes the convex hull and the multiplicity of $\lambda_{\max}(b^*)$ is m.

Even though the optimization problem (2) can be simplified to the minimization problem (4), we can not solve it as G is unknown. Hence, we follow the approach of Romer et al. (2018) and solve (2) by gradient-based methods.

3. SURROGATE MODELS FROM INPUT-OUTPUT DATA

In this section, we study gradient-based methods to solve (2). As already exploited in Wahlberg et al. (2010), the gradients

$$\begin{aligned}\nabla_u\rho =& \frac{2}{||u||^2}(PGPGu - PGPTu - T^TGu + T^TTu)\\ &- \frac{2}{||u||^4}u^T(PGPGu - PGPTu - T^TGu + T^TTu)u\\ \nabla_{b_i}\rho =& \frac{2}{||u||^2}u^T\tilde{T}_i^T(Tu - Gu) \end{aligned} \tag{5}$$

for all $(b,u) \in \mathbb{R}^N \times S^{n-1}$ can be concluded from three data samples with inputs u, PGu, and PTu. Here, P denotes the involutory permutation matrix

$$P = \begin{bmatrix} 0 & & 1\\ & \iddots & \\ 1 & & 0\end{bmatrix}.$$

Thus, we can solve (2) by algorithms from gradient-based optimization without knowledge of G. For the sake of space limitation, we only refer to robustness properties with respect to noise of general gradient-based algorithms, which are discussed in Koch et al. (2019) more thoroughly. Furthermore, if the output of the system is corrupted by additive white noise with zero mean, then the expected value of the noise-corrupted gradients retrieve the true gradients.

In the following section, we will study continuous-time saddle-point dynamics and examine their equilibrium points. Moreover, insight into the problem at hand, stability and convergence guarantees are provided. Subsequent, we will use these results to provide dedicated sampling schemes to find the 'best fitting' surrogate model.

3.1 Continuous-time saddle-point dynamics

The continuous-time saddle-point dynamics for the optimization problem (2) read

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}\tau}b_i(\tau) &= -\nabla_{b_i}\rho(b(\tau),u(\tau)),\ i=1,\dots,N\\ \frac{\mathrm{d}}{\mathrm{d}\tau}u(\tau) &= \nabla_u\rho(b(\tau),u(\tau)),\end{aligned} \tag{6}$$

with the gradients

$$\begin{aligned}\nabla_u\rho(b,u) &= \frac{2}{||u||^2}(A(b)-\rho(b,u)I)u\\ \nabla_{b_i}\rho(b,u) &= \frac{2}{||u||^2}u^T\tilde{T}_i^T(T(b)-G)u,\ i=1,\dots,N.\end{aligned}$$

As shown in Romer et al. (2018), the Euclidean norm of $u(\tau)$ is invariant under the saddle-point dynamics, and hence we consider the saddle-point dynamics on the unit sphere $u \in S^{n-1}$. Thus, the equilibria $(\bar{b},\bar{u})$ of the saddle-point dynamics (6) satisfy

$$\xi^T\nabla_u\rho(\bar{b},\bar{u}) = 0, \quad \forall\xi \in T_{\bar{u}}S^{n-1},$$

where $T_{\bar{u}}S^{n-1}$ denotes the tangent space on S^{n-1} at $\bar{u}$. By a simple calculation, this is equivalent to

$$A(\bar{b})\bar{u} = \lambda(\bar{b})\bar{u}, \tag{7}$$

which is also referred as the Courant-Fischer-Weyl principle. For that reason, the input $\bar{u}$ of any equilibrium of the saddle-point dynamics is an eigenvector of $A(\bar{b})$. This implies that any linear combination of eigenvectors of one eigenvalue is an equilibrium point if $\nabla_b\rho(\bar{b},\bar{u}) = 0$, which is equivalent to

$$\begin{bmatrix}\bar{u}^T\tilde{T}_1^T\tilde{T}_1\bar{u} & \cdots & \bar{u}^T\tilde{T}_1^T\tilde{T}_N\bar{u}\\ \vdots & \ddots & \vdots\\ \bar{u}^T\tilde{T}_N^T\tilde{T}_1\bar{u} & \cdots & \bar{u}^T\tilde{T}_N^T\tilde{T}_N\bar{u}\end{bmatrix}\bar{b} = \begin{bmatrix}\bar{u}^T\tilde{T}_1^TG\bar{u}\\ \vdots\\ \bar{u}^T\tilde{T}_N^TG\bar{u}\end{bmatrix} \tag{8}$$

by (5). Furthermore, (7) implies that the solution (b^*,u^*) of the optimization problem (2) is an equilibrium point of the saddle-point dynamics, as u^* lies in the eigenspace of $\lambda_{\max}(A(b^*))$. Moreover, we can show that (b^*,u^*) is a saddle-point

$$\rho(b^*,u) \le \rho(b^*,u^*) \le \rho(b,u^*),$$

as the second order gradients of $\rho(b,u)$ at (b^*,u^*) read

$$\begin{aligned}\nabla_{uu}\rho(b^*,u^*) &= 2(A(b^*) - \rho(b^*,u^*)I)\\ \nabla_{bb}\rho(b^*,u^*) &= 2\begin{bmatrix}{u^*}^T\tilde{T}_1^T\tilde{T}_1u^* & \dots & {u^*}^T\tilde{T}_1^T\tilde{T}_Nu^*\\ \vdots & \ddots & \vdots\\ {u^*}^T\tilde{T}_N^T\tilde{T}_1u^* & \dots & {u^*}^T\tilde{T}_N^T\tilde{T}_Nu^*\end{bmatrix}.\end{aligned}$$

Since $\rho(b^*,u^*)$ is equal to the maximal eigenvalue of $A(b^*)$, $\nabla_{uu}\rho(b^*,u^*)$ is negative semi-definite. Furthermore, the matrix $\nabla_{bb}\rho(b^*,u^*)$ is a Gramian matrix, and therefore positive semi-definite. Together with the knowledge that the first order gradients of $\rho(b,u)$ at (b^*,u^*) are zero, (b^*,u^*) is a saddle-point.

We have shown that the solution of (2) is an equilibrium point of the saddle-point dynamics. As in Romer et al. (2018), we will prove asymptotic stability of (b^*,u^*) under the saddle-point dynamics. Therefore, the saddle-point dynamics converge to the equilibrium (b^*,u^*) if the initialization lies in its neighbourhood.

In the literature, for example, Cherukuriy et al. (2015), stability results for saddle-points under saddle-point dynamics are mostly given for convex-concave saddle-points where $b \mapsto \rho(b,u)$ has to be convex and $u \mapsto \rho(b,u)$ has to be concave in the neighbourhood of the saddle-point. However, since the analysis of the concavity of $u \mapsto \rho(b,u)$ is not obvious, we will analyse the stability of (b^*,u^*) locally.

For this purpose, we introduce the Jacobian matrix J of the saddle-point dynamics

$$J(b,u) = \begin{bmatrix}-\nabla_{bb}\rho(b,u) & -\nabla_{bu}\rho(b,u)^T\\ \nabla_{bu}\rho(b,u) & \nabla_{uu}\rho(b,u)\end{bmatrix}.$$

Moreover, we define the i-th eigenvector w_i of $-\nabla_{bb}\rho(b^*,u^*)$ with $||w_i|| = 1$ and its corresponding eigenvalue $\lambda_{w_i} \le 0$. Furthermore, the i-th eigenvector of $A(b^*)$ is denoted by v_i with $||v_i|| = 1$ and $v_1 = u^*$. Since $\nabla_{bb}\rho(b^*,u^*)$ and $A(b^*)$ are symmetric, $w_i \in \mathbb{R}^N, i = 1,\dots,N$ and $v_i \in \mathbb{R}^n, i = 1,\dots,n$ span each an orthonormal basis.

Assumption 1. Let the Gramian matrix $\nabla_{bb}\rho(b^*, u^*)$ be positive definite.

Assumption 2. Let the vectors

$$\begin{bmatrix} w_1^T \nabla_{bu}\rho(b^*, u^*)^T v_2 \\ \vdots \\ w_N^T \nabla_{bu}\rho(b^*, u^*)^T v_2 \end{bmatrix}, \dots, \begin{bmatrix} w_1^T \nabla_{bu}\rho(b^*, u^*)^T v_m \\ \vdots \\ w_N^T \nabla_{bu}\rho(b^*, u^*)^T v_m \end{bmatrix}$$

be linearly independent.

Theorem 3. Suppose Assumption 1 and 2 hold. Moreover, let the eigenvalues of $A(b^*)$ be given by $\lambda_{\max}(b^*) = \lambda_1(b^*) = \cdots = \lambda_m(b^*) > \lambda_{m+1}(b^*) \geq \cdots \geq \lambda_n(b^*)$. Then the solution of the optimization problem (2) is exponentially stable under the saddle-point dynamics.

Proof. We show the claim by proving that the linearized saddle-point dynamics are asymptotically stable. Similar to Romer et al. (2018), we project the linearized dynamics on the tangent space of $\mathbb{R}^N \times S^{n-1}$ at (b^*, u^*) by

$$\begin{bmatrix} b \\ u \end{bmatrix} = \begin{bmatrix} W & 0 \\ 0 & V \end{bmatrix} x,$$

with $W = [w_1 \cdots w_N]$, $V = [v_2 \cdots v_n]$, and $x \in \mathbb{R}^{N+n-1}$. Thus, the transformed linearized system reads

$$\dot{x} = \begin{bmatrix} \lambda_{w_1} & & 0 & * & \dots & * \\ & \ddots & & \vdots & & \vdots \\ 0 & & \lambda_{w_N} & * & \dots & * \\ * & \dots & * & 2(\lambda_2 - \lambda_1) & & 0 \\ \vdots & & \vdots & & \ddots & \\ * & \dots & * & 0 & & 2(\lambda_n - \lambda_1) \end{bmatrix} x \quad (9)$$

$$=: \begin{bmatrix} \tilde{J}_{11} & -\tilde{J}_{12}^T \\ \tilde{J}_{12} & \tilde{J}_{22} \end{bmatrix} x =: \tilde{J}x.$$

To show asymptotic stability of $\dot{x} = \tilde{J}x$, we choose the Lyapunov function $V(x) = x^T x$. Since

$$\dot{V}(x) = x^T(\tilde{J} + \tilde{J}^T)x = 2x^T \begin{bmatrix} \tilde{J}_{11} & 0 \\ 0 & \tilde{J}_{22} \end{bmatrix} x \leq 0$$

for all $x \in \mathbb{R}^{n+N-1}$, we apply LaSalle's invariance principle to imply asymptotic stability. Under the assumption on the eigenvalues of $A(b^*)$ and Assumption 1, the set of states with $\dot{V}(x) = 0$ is given by

$$E := \{x = \begin{bmatrix} 0_N^T & \alpha_2 & \cdots & \alpha_m & 0_{n-m}^T \end{bmatrix}^T, \alpha_i \in \mathbb{R}\},$$

where 0_N denotes the zero vector of length N. Hence, the system dynamics in E yield

$$\tilde{J} \begin{bmatrix} 0_N \\ \alpha_2 \\ \vdots \\ \alpha_m \\ 0_{n-m} \end{bmatrix} = - \begin{bmatrix} \tilde{J}_{12}^T \begin{bmatrix} \alpha_2 \\ \vdots \\ \alpha_m \\ 0_{n-m} \end{bmatrix} \\ 0_{n-1} \end{bmatrix}. \quad (10)$$

Since the first $m-1$ columns of $\tilde{J}_{12}^T$ are linearly independent by Assumption 2, the first N entries of $\dot{x}$ are unequal to zero for $[\alpha_2 \dots \alpha_m] \neq 0$. Thereby, the system (9) stays in E only for $\alpha_2 = \cdots = \alpha_m = 0$, and hence the largest invariant subset of E is $x = 0$. By the LaSalle's invariance principle, $\dot{x} = \tilde{J}x$ is asymptotically stable, and therefore the saddle-point dynamics are exponentially stable at (b^*, u^*). □

In the sequel, we comment on the three assumptions required in Theorem 3 and show that these assumptions hold for almost all systems G and choices of $\tilde{T}_i$. First, since the case $m > 2$ is fairly rare, Assumption 2 reduces to

$$\begin{bmatrix} w_1^T \nabla_{bu}\rho(b^*, u^*)^T v_2 \\ \vdots \\ w_N^T \nabla_{bu}\rho(b^*, u^*)^T v_2 \end{bmatrix} \neq 0.$$

In equation (10), Assumption 2 is exploited to apply LaSalle's invariance principle, and hence to imply attractivity of the saddle-point. Thus, Assumption 2 guarantees that (b^*, u^*) is an isolated equilibrium. Note that, another geometric interpretation of Assumption 2 is discussed in Koch et al. (2019) for the case $N = 1$.

Second, Assumption 1 is satisfied if and only if $\tilde{T}_i u^*, i = 1, \dots, N$ are linearly independent. With $n \gg N$, this constitutes a quite mild assumption. However, even if $\nabla_{bb}\rho(b^*, u^*)$ is just positive semi-definite, then exponential stability of (b^*, u^*) can be proven under an additional rank condition on $\tilde{J}_{12}$. For the sake of place limitation, we skip this proof as he follows the proof of Theorem 3.

Note that, Theorem 3 implies that the saddle-point dynamics converge to (b^*, u^*) only if the initialization lies in its neighbourhood. However, even though the optimal surrogate model may not be found by the saddle-point dynamics, the bound Δ of the maximal error is guaranteed for the suboptimal model $T(b)$ for any $b \neq b^*$, if the saddle-point dynamics converge to an equilibrium point in the eigenspace of the largest eigenvalue of $A(b)$.

3.2 Iterative methods to find the 'optimal' surrogate model

In the previous subsection, we examined the continuous-time saddle-point dynamics to find the solution (b^*, u^*) of the optimization problem (2). However, the gradients can only be evaluated by experiments. Thus, it is necessary to solve the optimization problem iteratively by discrete-time optimization. In the following, two iterative schemes from Arrow et al. (1958) and Polyak (1970) are applied to find the saddle-point (b^*, u^*) of $\rho(b, u)$.

First, the Arrow-Hurwicz iteration was introduced in Arrow et al. (1958), which corresponds to a time discretization of the saddle-point dynamics (6) with step size α

$$\begin{aligned} b(k+1) &= b(k) - \alpha\nabla_b\rho(b(k), u(k)) \\ u(k+1) &= u(k) + \alpha\nabla_u\rho(b(k), u(k)). \end{aligned} \quad (11)$$

Whereas Romer et al. (2018) presents a modified Arrow-Hurwicz iteration

$$\begin{aligned} b(k+1) &= b(k) - \alpha\nabla_b\rho(b'(k), u'(k)) \\ u(k+1) &= \frac{u(k) + \alpha\nabla_u\rho(b'(k), u'(k))}{||u(k) + \alpha\nabla_u\rho(b'(k), u'(k))||} \\ b'(k+1) &= b(k+1) - \alpha\nabla_b\rho(b'(k), u'(k)) \\ u'(k+1) &= \frac{u(k+1) + \alpha\nabla_u\rho(b'(k), u'(k))}{||u(k+1) + \alpha\nabla_u\rho(b'(k), u'(k))||}, \end{aligned} \quad (12)$$

which ensures that the sequence $u(k)$ stays on S^{n-1} for all $k \in \mathbb{N}$. In Koch et al. (2019), a modified Uzawa iteration was shown

$$\begin{aligned} b(k+1) &= \underset{b \in \mathbb{R}^N}{\operatorname{argmin}}\ \rho(b, u(k)) \\ u'(k+1) &= u(k) + \alpha\nabla_u\rho(b(k), u(k)) \\ u(k+1) &= \frac{u'(k+1)}{||u'(k+1)||}, \end{aligned} \quad (13)$$

where $||u(k)|| = 1$ for all $k \in \mathbb{N}$ is ensured. In the Uzawa iteration, the gradient descent step of the Arrow-Hurwicz iteration is replaced by minimizing $\rho(b, u(k))$ with respect to b in order to potentially increase its convergence rate. By the first order necessary condition (8), the solution of this minimization is given by $C(u(k))b(k+1) = d(u(k))$, with

$$C(u) = \begin{bmatrix} u^T \tilde{T}_1^T \tilde{T}_1 u & \dots & u^T \tilde{T}_1^T \tilde{T}_N u \\ \vdots & \ddots & \vdots \\ u^T \tilde{T}_N^T \tilde{T}_1 u & \dots & u^T \tilde{T}_N^T \tilde{T}_N u \end{bmatrix}$$

$$d(u) = \begin{bmatrix} u^T \tilde{T}_1^T G u \\ \vdots \\ u^T \tilde{T}_N^T G u \end{bmatrix}.$$

Hence, $b(k+1)$ is computable under the knowledge of the input-output tuple $(u(k), Gu(k))$. Note that, the Hankel-matrix $C(u)$ might be positive semi-definite, and thus $b(k+1)$ is not unique. To avoid large oscillation of the sequence $b(\cdot)$, we choose $b(k+1)$ in the set of solutions which is closest to $b(k)$. Note that, all $b(k+1)$ in the set of solutions are global minima, as the optimization problem

$$b(k+1) = \underset{b \in \mathbb{R}^N}{\operatorname{argmin}}\ \rho(b, u(k))$$

is convex.

We finish this section by providing some convergence results from Polyak (1970) for the Arrow-Hurwicz iteration and Uzawa iteration.

Proposition 4. Let the eigenvalues of $A(b^*)$ be given by $\lambda_{\max}(b^*) = \lambda_1(b^*) = \dots = \lambda_m(b^*) > \lambda_{m+1}(b^*) \geq \dots \geq \lambda_n(b^*)$. Under the Assumption 1 and Assumption 2, there exists an $\bar{\alpha}$ such that the Arrow-Hurwicz iteration (11) is locally convergent to (b^*, u^*) for $0 < \alpha < \bar{\alpha}$.

Proof. The proof is based on the linearization of the Arrow-Hurwicz iteration and exploiting that the projected Jacobian matrix $\tilde{J}$ on the tangent space of $\mathbb{R}^N \times S^{n-1}$ at (b^*, u^*) is Hurwicz. With the results from Theorem 3, the proof in Romer et al. (2018) can be adjusted accordingly.

A similar result is discussed for the modified Arrow-Hurwicz iteration (12) in Romer et al. (2018).

Proposition 5. Let the eigenvalues of $A(b^*)$ be given by $\lambda_{\max}(b^*) = \lambda_1(b^*) = \dots = \lambda_m(b^*) > \lambda_{m+1}(b^*) \geq \dots \geq \lambda_n(b^*)$. Under the Assumption 1 and Assumption 2, there exists an $\tilde{\alpha}$ such that the Uzawa iteration (13) is locally convergent to (b^*, u^*) for $0 < \alpha < \tilde{\alpha}$.

Proof. We refer to the proof in Polyak (1970), that can be adapted with the results from Theorem 3.

4. NUMERICAL EXAMPLE

In this section, we will apply the modified Arrow-Hurwicz iteration (12) and the Uzawa iteration (13) to find the 'best fitting' surrogate model for a LTI SISO benchmark system for model order reduction methods from Chahlaoui and Dooren (2005). The system *beam* models a clamped beam, where the input represents a force applied at the free end and the output is the resulting displacement. The system contains 348 states, which result from the discretization of a partial differential equation. We use an exact time discretization with zero-order hold and time step 10. Each experiment has a length of $n = 100$ time steps. The initial input is chosen as a normalized step signal $u(0) = 10^{-1}\,[1\ \cdots\ 1]^T$ and the initial model parameter as $b(0) = [1\ \cdots\ 1]^T$.

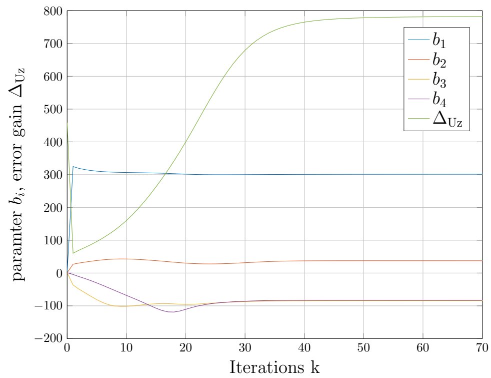

Fig. 3. Simulation results for the Uzawa iteration.

For this example, we expect an oscillating input-output behaviour. Hence, the surrogate model is chosen as a sum of stable LTI systems of order two

$$y(t) = \sum_{i=1}^{N} \frac{b_i q^{-1}}{1 - a_{1i} q^{-1} - a_{2i} q^{-2}} u(t). \tag{14}$$

While the parameter b_i are optimized by our schemes, the parameter $a_{1i} \in \{1, 1.7, 0.3, -0.06, \dots\}$ and $a_{2i} = -0.92$ are chosen such that the poles of (14) are complex conjugated and stable. Note that, we considered the step response of the clamped beam to find a suitable choice of a_{1i} and a_{2i}. The step sizes for the modified Arrow-Hurwicz iteration and the Uzawa iteration are chosen to $\alpha_{\mathrm{AH}} = 6 \cdot 10^{-7}$ and $\alpha_{\mathrm{Uz}} = 10^{-7}$, respectively. The results of the simulations are shown in the following table.

Table 1. Bound Δ of the maximal error and parameters b for increasing size of the surrogate model.

	N=1	N=2	N=4
Δ^*	1004	948	785
Δ_{AH}	1004	948	786
Δ_{Uz}	1004	948	786
b^*	305.5	$[307.1\ \ 46.1]$	$[302.2\ \ 37.5\ \ {-83.7}\ \ {-82.7}]$
b_{AH}	305.7	$[307.1\ \ 46.2]$	$[302.1\ \ 37.5\ \ {-83.9}\ \ {-82.4}]$
b_{Uz}	305.6	$[307.1\ \ 46.2]$	$[302.0\ \ 37.4\ \ {-83.9}\ \ {-82.9}]$

Note that, the optimal parameters b^* are calculated under the knowledge of G by considering the maximal eigenvalue function $\lambda_{\max}(b)$ and solving the minimization problem (4). However, this is only possible for $N \leq 4$ as the computation time increases exponentially. Table 1 shows, that the Arrow-Hurwicz iteration and the Uzawa iteration converge to the solution of the optimization problem (b^*, u^*). However, as shown in Fig. 3 and Fig. 4, the number of iterations for convergence of the Arrow-Hurwicz iteration is significantly higher than of the Uzawa iteration. Hence, the Uzawa iteration (13) exhibits a faster convergence than the Arrow-Hurwicz iteration in this numerical example.

As expected, the bound Δ of the maximal error decreases by increasing the number of basis systems N. For $N = 8$, Δ can be decreased to 582, while the number of iterations for convergence increases to 150. In this case, even though we cannot guarantee that the optimal parameter vector

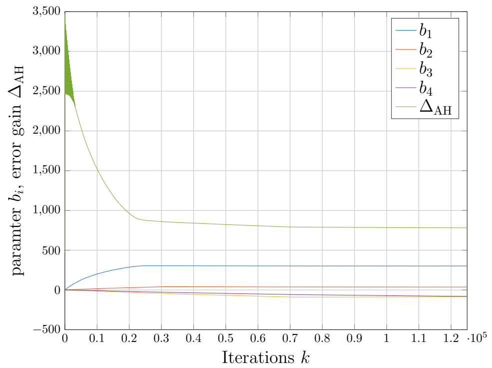

Fig. 4. Simulation results for the Arrow-Hurwicz iteration.

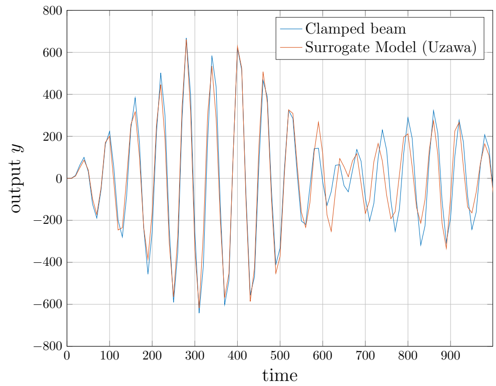

Fig. 5. Output of the optimized surrogate model and the system for the 'worst-case input'.

is found, the bound on the error actually holds, as the sequence $u(\cdot)$ converges to the eigenspace of the largest eigenvalue of $A(b(k))$ for $k \geq 200$. However, for this example, the output of the optimized surrogate model almost coincides with the output of the system already for $N = 4$ for the 'worst-case input' as shown in Fig. 5. This example also shows that we can improve the controller design using the obtained surrogate model compared to a controller design via the small-gain theorem or concic relations, respectively, in the sense mentioned in Section 2. Indeed, the norm $||G||_\infty$ with 3864 and the radius of the smallest cone with 3701 are significantly higher than Δ^*.

5. CONCLUSION

In this paper, we proposed a data-driven approach to iteratively compute a low-order surrogate model for large-scale LTI SISO systems and its guaranteed smallest upper bound of the approximation error. To solve the corresponding optimization problem, gradient-based methods from optimization were applied and their convergence behaviour were analysed.

In the proposed schemes, we were restricted to surrogate models which are linear combination of LTI systems. This was important, among other reasons, to apply the Uzawa iteration without knowledge of the system's input-ouput behaviour. Thus, the linearity of the surrogate model in the parameter was the key to keep the number of iterations small. However, since finding suitable basis systems can be difficult, it might be interesting to develop schemes, where the surrogate model is nonlinear in the parameters. Moreover, extending the idea to nonlinear systems might be interesting to compute the nonlinearity measure of a system, which are investigated in Martin and Allgöwer (2019) for data-driven system analysis and controller design.

REFERENCES

Antoulas, A.C. (2005). *Approximation of Large-Scale Dynamical Systems.* Advances in Design and Control, SIAM.

Arrow, K.J., Hurwicz, L. and Uzawa H. (1958). *Studies in linear and non-linear programming.* Stanford University Press.

Chahlaoui, Y. and van Dooren, P. (2005). Benchmark Examples for Model Reduction of Linear Time-Invariant Dynamical Systems. In *Dimension Reduction of Large-Scale Systems*, 45, 379-392, Springer.

Cherukuri, A., Gharesifard, B. and Cortés, J. (2015). Saddle-Point Dynamics: Conditions for Asymptotic Stability of Saddle Points. *SIAM J. Control and Optimization*, 55(1), 486-511.

Feijer, D. and Paganini F. (2010). Stability of primal–dual gradient dynamics and applications to network optimization. *Automatica*, 46(12), 1974–1981.

Ionita, A.C. and Antoulas, A.C. (2014). Data-Driven Parametrized Model Reduction in the Loewner Framework. *SIAM J. Sci. Comput.*, 36(3), A984-A1007.

Martin, T. and Allgöwer, F. (2019). Nonlinearity measures for data-driven system analysis and control. In *Proc. 58th Conf. on Decision and Control* .

Mengi, E., Yildirim, E.A. and Kilic, M. (2014). Numerical Optimization of Eigenvalues of Hermitian Matrix Functions. *SIAM J. Matrix Anal. Appl.*, 35(2), 699-724.

Milanese, M. and Novara, C. (2004). Set Membership identification of nonlinear systems. *Automatica*, 40(6), 957–975.

Oomen, T., van der Maas, R., Rojas, C.R. and Hjalmarsson, H. (2014). Iterative Data-Driven $\mathcal{H}_\infty$ Norm Estimation of Multivariable Systems With Application to Robust Active Vibration Isolation. *IEEE Trans. Ctrl. Sys. Tech.*, 22(6), 2247-2260.

Polyak, B.T. (1970). Iterative methods using lagrange multipliers for solving extremal problems with constraints of the equation type. *USSR Comput. Math. and Math. Phys.*, 10(5), 42-52.

Popov, L.D. (1980). A modification of the Arrow-Hurwicz method for search of saddle points. *Mat. Zametki*, 28(5), 777-784.

Romer, A., Montenbruck, J.M. and Allgöwer, F. (2018). Data-driven inference of conic relations via saddle-point dynamics. In *Proc. 9th IFAC Symp. Robust Control Design*, 51(25), 396-401.

Koch, A., Montenbruck, J.M. and Allgöwer, F. (2019). Sampling Strategies for Data-Driven Inference of Input-Output System Properties. *arXiv preprint* arXiv: 1910.08919.

Scarciotti, G. and Astolfi, A. (2017). Data-Driven model reduction by moment matching for linear and nonlinear systems. *Automatica*, 79, 340-351.

Wahlberg, B., Syberg, M.B. and Hjalmarsson, H. (2010). Non-parametric methods for $\mathcal{L}_2$-gain estimation using iterative experiments. *Automatica*, 46, 1376-1381.

Zames, G. (1966). On the input-output stability of time-varying nonlinear feedback systems part I: Conditions derived using concepts of loop gain, conicity, and positivity. *IEEE, Trans. Automat. Control*, 11, 228-238.