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

Linearization in the Large of the Anaerobic Digestion Process Using a Reduced-Order Koopman Operator

Camilo Garcia-Tenorio [*,**] **Mihaela Sbarciog** [**]
Eduardo Mojica-Nava [*] **Alain Vande Wouwer** [**]

[*] *Universidad Nacional, Bogotá, Colombia (e-mail: (cagarciate, eamojican)@unal.edu.co)*
[**] *Université de Mons, Mons, Belgique (e-mail: mihaela@sbarciog.be, alain.vandewouwer@umons.ac.be)*

Abstract: The identification of accurate models for the anaerobic digestion process is essential for the characterization of the region that guarantees the conservation of the bacteria population. Traditional techniques involve the identification of nonlinear models based on data from the system. In this paper, we introduce data-driven techniques that allow the characterization of the system's behavior via the approximation of the Koopman operator with the extended dynamic mode decomposition algorithm. We propose methods to reduce the order and dimension of the representation based on orthogonal polynomials.

Keywords: Anaerobic Digestion, Eigenfunction, Eigenvalue, EDMD, Koopman Operator, Nonlinear Systems.

1. INTRODUCTION

The biological process through which multiple organisms break down organic matter in the absence of oxygen is called Anaerobic Digestion (AD). What motivates the use of the AD process on an industrial scale is its capacity of degrading strong and resilient substrates, the low sludge production, and the possibility of making a profit out of the formation of methane gas (Mailleret et al., 2003).

However, the AD process is difficult to operate and control. Knowledge and expertise are necessary to maintain the stability of the process and avoid washout conditions. Varying operating variables such as the concentration of organic matter at the input feed, the bioreactor dilution rate, or an accumulation of intermediate compounds can destabilize the process (Sbarciog et al., 2010).

To study the dynamics of this process, there are well-accepted and detailed models, such as the Anaerobic Digestion Model 1 (ADM1) (Batstone et al., 2002). Although the customization of this model fits a variety of wastes and plant configurations, its high order with 32 state variables and general complexity raise issues for controller analysis and design (Giovannini et al., 2018). There are low-order models to overcome this issue, such as the Anaerobic Model 2 (AM2) (Bernard et al., 2001), which remains highly nonlinear. These systems give a proper description of the dynamics, but the task of calibrating the model from experimental data using identification techniques is difficult ((Donoso-Bravo et al., 2011; Yu and Wensel, 2013) and the references therein).

⋆ Camilo Garcia-Tenorio is supported by Colciencias-Doctorado Nacional-647/2015

The idea of this work is to develop low-order/dimensional linear models of the process in a data-driven approach based on selected measurement signals, such as dilution rate, biomass or substrate concentrations. The chosen method is the Extended Dynamic Mode Decomposition (EDMD) algorithm (Williams et al., 2016a). This algorithm gives a truncated approximation of the Koopman operator, where a set of functions Ψ of a linear operator U are observables or measures of a particular discrete-time nonlinear dynamical system ($x(k+1) = T(x(k))$), i.e., the evolution of these functions relate to the evolution of the states of the system (Koopman, 1931). The application of U on Ψ gives the same result as applying the observables Ψ on the state evolution

$$U\Psi(x) = \Psi(T(x))$$

This approach allows for an analysis of the evolution of functions rather than the analysis of the evolution of states (Budišić et al., 2012). The trade-off with this approach concerns linearity and dimensionality, as a finite-dimensional nonlinear system will be described by an infinite-dimensional linear system.

The proposed method to get a low-order/dimensional approximation of the Koopman operator for the AD process is twofold. First, a p-q-quasi norm reduction method, based on the proposed ideas for reliability analysis in polynomial chaos (Sudret, 2001), reduces the maximum order and the dimension of the observable functions. Second, an original method is proposed based on an error criterion for the individual elements of the observable function, which serves to identify the elements that could be eliminated from the basis, thus, reducing the dimension even further.

The rest of this paper is organized as follows. Section 2 gives the general description of the AD process, dynamics, chosen behavior of the system, and the general objectives. Sections 3 and 4 provide the current and proposed methods to get the approximation of the Koopman operator, and the low-order/dimensional approximation, and Section 5 presents the results of the methodology applied to the AD process.

Notation: $\mathbb{C}$ denotes the field of complex numbers. $\mathbb{R}$ and $\mathbb{R}_0^+$ denote the field of real and nonnegative real numbers, respectively. For any matrix $A \in \mathbb{R}^{n\times n}$, $A^\star$ denotes complex conjugate transpose, A^+ denotes its pseudoinverse, and $\|x\|$ represents the Euclidean norm. The space $n \times n$ of real matrices is denoted by $R^{n\times n}$. For a complex number λ, $|\lambda|$ represents its norm. For any set A, $\bar{A}$ denotes the closure of A. The operator $\circ$ is defined as the product term to term. The vector exponentiation $M^{\pm\eta}$ is defined term by term. A level set of an arbitrary function $h(x)$ for any constant c is $\Gamma(h(x)) = \{x \in \mathbb{R}^n : h(x) = c\}$.

2. ANAEROBIC DIGESTION SYSTEM

Anaerobic digestion of the effluent waste from winery and brewery production is an important downstream process. In addition to wastewater treatment, anaerobic digestion can also provide biogas that can be used as a source of renewable energy. However, this process is delicate to operate and to start up, particularly because (Dochain et al., 1991):

(i) The process is very unstable in the sense that variations in the hydraulic flow at the input, as well as the concentration of organic load can lead the bioreactor into the wash out state where the living bacterial population disappears.
(ii) The output pollution has to be maintained at a prescribed level despite the fluctuations in the organic load at the input.

These factors are difficult to master because the anaerobic digestion process involves several bacterial populations, which interact in a nonlinear fashion. In a simplified view (considering only two main populations), the interaction between the acid-forming bacteria, and the methane forming bacteria can produce an imbalance leading to the washout of one or both of the populations. The dynamics of this simplified view, where there is only two reactions: acidogenesis and methanogenesis is described by the following reaction network:

$$\begin{aligned} a\xi_1 &\xrightarrow{r_1(\xi)} c\xi_2 + \xi_3, \\ d\xi_2 &\xrightarrow{r_2(\xi)} \mathrm{CH}_4 + \xi_4. \end{aligned} \tag{1}$$

In the first reaction, the acidogenic bacteria ξ_3 consumes the organic substrate ξ_1 for growth and produces volatile fatty acids ξ_2. In the second reaction, the methanogenic bacteria ξ_4 uses the volatile fatty acids as a substrate for growth and produces methane. It is necessary to maintain the balance between the acidogenesis and the methanogenesis states in the operation of the AD process to avoid acidification, which is the state of accumulation of volatile fatty acids in the reactor that produces the washout of methanogenic bacteria.

For an ideal continuous stirred tank reactor, the following differential equation describes the system dynamics of the reaction network (1):

$$\begin{aligned} \dot{\xi}_1 &= u(\xi_{\mathrm{in}_1} - \xi_1) - ar_1(\xi), \\ \dot{\xi}_2 &= u(\xi_{\mathrm{in}_2} - \xi_2) + cr_1(\xi) - dr_2(\xi), \\ \dot{\xi}_2 &= -u\xi_3 + r_1(\xi), \\ \dot{\xi}_4 &= -u\xi_4 + r_2(\xi). \end{aligned}$$

Where $\xi = [\xi_1\ \xi_2\ \xi_3\ \xi_4]^T \in \mathbb{R}_+^4$ is the state vector, $u \in \mathbb{R}_+$ is the dillution rate, $\xi_{\mathrm{in}_1}, \xi_{\mathrm{in}_2} \in \mathbb{R}_+$ are the concentrations of organic substrate and volatile fatty acids in the influent, $a, c, d \in \mathbb{R}_+$ are the stoichiometric coefficients, and $r_1(\xi), r_2(\xi)$ are the reaction rates defined as

$$\begin{aligned} r_1(\xi) &= f_1(\xi_1)\xi_3, \\ r_2(\xi) &= f_2(\xi_2)\xi_4, \end{aligned}$$

where the the growth functions, $f_1(\xi_1)$ and $f_2(\xi_2)$ are based on Monod and Haldane type respectively as

$$\begin{aligned} f_1(\xi_1) &= \mu_{m_1}\frac{\xi_1}{K_{s_1} + \xi_1}, \\ f_2(\xi_2) &= \mu_{m_2}\frac{\xi_2}{K_{s_2} + \xi_2 + \frac{\xi_2^2}{K_{i_2}}}. \end{aligned}$$

A canonical state space transformation of the AD system can be obtained by considering the partition $\xi = [\xi_a\ \xi_b]^T$, where $\xi_a = [\xi_3\ \xi_4]^T$ and $\xi_b = [\xi_1\ \xi_2]^T$, and a linear transformation of the states $x_a = \xi_a$ and $x_b = \xi_b - C_bC_a^{-1}\xi_a$ Bastin and Dochain (1990); Bastin and Van Impe (1995) as,

$$\begin{aligned} \dot{x}_a &= u(w_a - x_a) + C_a\rho(x), \\ \dot{x}_b &= u(w_b - x_b), \end{aligned} \tag{2}$$

with $x_a \triangleq [x_3\ x_4]^T$ are the state variables that represent the acidogenic and methanogenic bacteria respectively, $x_b \triangleq [x_1\ \ x_2]^T$, are the state variables that represent the linear combination of substrates and species, $C_a = I_2$, $\rho(x) \triangleq [\rho_1(x)\ \rho_2(x)]$ are the reaction rates, $w_a \triangleq [w_3\ w_4]^T = [0\ 0]^T$ are the concentration of species in the input flow, $w_b \triangleq [w_1\ w_2]^T = [\xi_{\mathrm{in}_1}\ \xi_{\mathrm{in}_2}]$ are the concentration of substrates in the input flow,

$$C_b = \begin{bmatrix} -a & 0 \\ c & -d \end{bmatrix},$$

is the stoichiometric matrix, and the reaction rates in the canonical state space are

$$\rho_i(x) = r_i(\xi)|_{\xi_a = x_a;\ \xi_b = x_b + C_bC_a^{-1}x_a},\ i = 1, 2.$$

Depending on the magnitude of the dilution rate and substrate concentration at the inflow, the system may possess up to six equilibrium points. In Sbarciog et al. (2010), the authors present the regions in the input space for which various numbers of equilibria occur and their stability.

This paper will consider the case in which six equilibrium points are present, where two of them are asymptotically stable (AS), one representing the desired working point of the system, where the acidogenic and methanogenic bacteria coexist, and the other, the acidification point. Figure 1 depicts the behavior of the nonlinear part of the system (2).

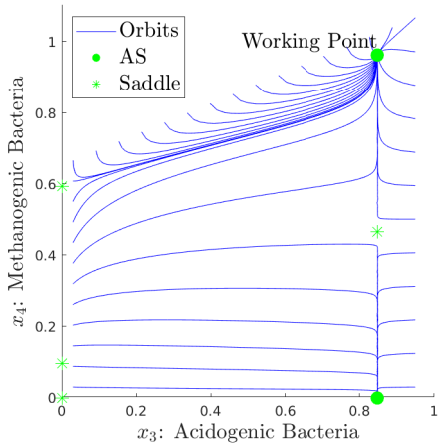

Fig. 1. Orbits and equilibrium points of the AD process.

2.1 Practical Considerations and objectives

Equation (2) has two distinct parts, a nonlinear dynamical system of the species x_a and a linear one representing a linear combination of the species and substrate x_b. Assuming that the linear part is in steady-state allows for the analysis of the system as a two-state variables system. This paper deals with this particular case, where there are measurements of the species concentrations.

In general, these measurements are not available. They often come from computer sensors from models such as (Bernard et al., 2001) whose identification comes from the available data such as gas flow-rate (sum of all gas compounds), the content of volatile suspended solids, volatile fatty acids, total organic carbon or chemical oxygen demand among others.

In this paper, the main objective is to establish the feasibility to get accurate models for the later approximation of the system based on available data. These models not only give an expanded linear representation of the system, but they also provide tools from the spectral analysis to obtain useful information such as the regions of attraction of the asymptotically stable equilibrium points (Garcia-Tenorio et al., 2019).

3. METHODS

This section describes the general theory on the Koopman operator for unforced systems and the current approach for the approximation of the operator via the EDMD algorithm for the forced case.

3.1 The Koopman Operator

Consider the autonomous nonlinear discrete-time dynamical system $(\mathcal{M}; T(x); k)$, with state variables $x \in \mathcal{M}$ where $\mathcal{M} \subseteq \mathbb{R}^n$ is the nonempty compact space state, $k \in \mathbb{Z}_0^+$ is the discrete time, and $T\colon \mathcal{M} \to \mathcal{M}$ is the differentiable vector-valued evolution map, i.e.,

$$x(k+1) = T(x(k)), \quad x_0 = x(0). \tag{3}$$

The solution to (3) is the successive application of T from an initial condition $x_0 \in M$ at $k = 0$, i.e., $x_k = T^k(x_0)$. This $x_k \in \mathcal{M}$ is an infinite sequence called a trajectory or orbit of the system.

For this discrete time system, consider a set of observation functions $f(x)\colon \mathcal{M} \to \mathbb{C}$ that belong to some function space, i.e., $f(x) \in \mathcal{F}$. For these functions there is an operator U^k, where the action of this operator on observables defines their evolution. This is the Koopman operator, and its relationship with the observable functions and the states of the system is

$$\left[U^k f\right](x) = f\left(T^k(x)\right).$$

In other words, the time-evolution of observations is the observations of the time-evolution of states. The trade-off with this approach concerns linearity and dimensionality; finite-dimensional nonlinear systems described by an infinite-dimensional linear one.

3.2 Extended Dynamic Mode Decomposition

Consider the autonomous nonlinear discrete dynamical system (3), the EDMD algorithm to approximate the Koopman operator in discrete time requires N pairs of data snapshots, either from a real system or a computationally integrated one at a specific sampling Δt. The snapshot pairs, $\{(x_i, y_i)\}_{i=1}^N$, where $y_i = T(x_i)$, are organized in data sets

$$X = [x_1\ x_2\ \dots\ x_N],\ \ Y = [y_1\ y_2\ \dots\ y_N].$$

The "extended" part of the EDMD algorithm consists in the approximation of the Koopman operator of a "lifted" space of the state variables, rather than approximating the space state as in the dynamic mode decomposition algorithm Schmid (2010). The "lifting" procedure consists in evaluating the space state of the system with a vector-valued function of observables $\Psi\colon \mathcal{M} \to \mathbb{C}^{d\times 1}$; $\Psi(x) = [\psi_1(x), \cdots, \psi_d(x)]^\top$.

The truncated linear operator $U_d \in \mathbb{C}^{d\times d}$ satisfies the following relation for the evaluation of the vector-valued function $\Psi(x)$ in the snapshot data of the system:

$$\Psi(Y) = U_d \Psi(X) + r(X),$$

where $r(X) \in \mathcal{F}$ is the residual term to minimize in order to find U_d. This minimization accepts a close solution within the least mean squares problem, where the objective function has the form

$$r(X) = \frac{1}{N}\sum_{i=1}^{N} \frac{1}{2} \left\| \Psi(y_i) - U_d \Psi(x_i) \right\|_2^2,$$

and the solution is

$$U_d \triangleq AG^+, \tag{4}$$

with matrices $G, A \in \mathbb{C}^{d\times d}$ defined by

$$G = \frac{1}{N}\sum_{i=1}^{N} \Psi(x_i)\Psi(x_i)^\top, \quad A = \frac{1}{N}\sum_{i=1}^{N} \Psi(x_i)\Psi(y_i)^\top.$$

The choice of the observables $\psi_l(x)$ usually consists of an orthogonal basis of polynomials Koekoek et al. (2010), radial basis functions Korda and Mezić (2018), or an arbitrarily constructed set with polynomial elements, and trigonometric functions, among others Brunton et al. (2016). The choice of orthogonal polynomials has the advantage of being able to recover the state linearly without

the need to add more elements to the observables function or by the solution of additional optimization problems.

The solution (4) gives a linear evolution of observables of the space state,

$$\Psi(x(k+1)) = U_d \Psi(x(k)), \tag{5}$$

and does not give the evolution of states. If the observables functions contain a direct observation of the state, i.e., $\psi_i(x) = x_i$ for $i = 1, \cdots, n$, the value of the states is directly available from the observables. Otherwise, if the functions come from a set of radial basis functions or other norm-based functions, the application of additional optimization solutions serve to recover the state. Or if the functions come from a set of orthogonal polynomials, the basis contains at least n injective functions of order one, such that the inverse of these functions gives the value of the state.

The topic of the next section is on how to select the polynomial elements of the observables function, and the proposal of error metrics to eliminate them to get a low order approximation of the linear operator that captures the dynamics of the system.

4. LOW-ORDER POLYNOMIAL BASIS

A sequence of orthogonal polynomials $\{\pi_\alpha(x)\}_{\alpha=0}^{\infty}$ where $\pi_\alpha(x)$ is a univariate (i.e., $x \in \mathbb{R}$) polynomial of degree α, is a sequence defined over a range $[a, b]$ where some inner product between distinct elements is zero, i.e., $\langle \pi_i(x), \pi_j(x) \rangle = 0$ for $i \neq j$. For the particular case of the approximation of the Koopman operator, either a sequence of Hermite polynomials or Laguerre polynomials is suitable for the approximation, as these are defined over the ranges $[-\infty, \infty]$ and $[0, \infty]$ respectively.

Every element of the multivariate vector-valued function of observables $\Psi(x)$ where $x \in \mathcal{M}$ and $\mathcal{M} \subseteq \mathbb{R}^n$ comes from the tensor product of n univariate polynomials as in

$$\psi_l(x) = \prod_{j=1}^{n} \pi_{\alpha_j}(x_j). \tag{6}$$

Consider for example a polynomial basis where $n = 2$ and a maximum degree of 2 for every univariate polynomial. The set of indices α for the construction of the polynomial basis is

$$\alpha = \begin{Bmatrix} \alpha_1 \\ \alpha_2 \end{Bmatrix} = \begin{Bmatrix} 0 & 1 & 2 & 0 & 1 & 2 & 0 & 1 & 2 \\ 0 & 0 & 0 & 1 & 1 & 1 & 2 & 2 & 2 \end{Bmatrix}.$$

Even though the maximum degree of the univariate polynomials is 2, the maximum degree of the multivariate elements is 4. By taking the full set of indices for every state variable, the dimension of the observables function is

$$\dim \Psi(x) = \binom{n+p+1}{p+1},$$

where p is the maximum degree of the univariate polynomials. This selection of indices poses a problem, as the dimension of the observables function grows exponentially with the addition of state variables or an increase in the maximum univariate degree. The growth of the dimensionality hinders the possibility of calculating the approximation of the Koopman operator due to the course of dimensionality problem and because of the error induced by high order elements in the polynomial basis.

To solve these problems, the use of p-q-quasi [1] norms (Konakli and Sudret (2016)) reduces the dimension of the basis by applying a truncation scheme on the maximum order of every multivariate element. This truncation scheme, for the set of indices of every element of the multivariate polynomial (6) is

$$\alpha_l = \{\alpha \in \mathbb{N}^n : \|\alpha\|_q \leq p\} \tag{7}$$

Consider the las example where $n = 2$ and apply the truncation scheme (7) with a maximum order $p = 3$ for the multivariate polynomial elements, and a norm $q = 0.7$. The new set of indices is

$$\alpha = \begin{Bmatrix} 0 & 1 & 2 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 & 1 & 2 \end{Bmatrix},$$

this truncation reduces the dimension of the vector-valued function from 8 to 6. The effect of this truncation is appreciated when n and p start to grow. For example, for the case of $n = 6$ and $p = 4$, considering a norm $q = 1$, the truncation scheme reduces the dimension of the vector-valued function from 4096 to 204.

Even though the p-q-quasi norm truncation scheme reduces the dimension of the vector-valued function, some of the multivariate polynomial elements do not contribute to the reduction of the error (1). The elimination of these elements reduces the dimension of the vector-valued function and, consequently, reduces the order of the calculated operator. The next section of this paper shows an error and elimination criteria for multivariate elements that achieve an additional reduction.

4.1 Reduction by Polynomial Accuracy

The idea of the reduction is to calculate the error of the individual multivariate polynomials and eliminate the elements whose error is greater than a specified threshold.

The approximation of the Koopman operator from the EDMD algorithm comes from a set of the system's orbits. This set of orbits comes from one of two distinct sets, one for solving the least-squares problem (9) and one for testing the accuracy of the solution. Evaluating the posterior time event Y in every element of the vector-valued function, and comparing it with the effect of the operator on the evaluation of the anterior time event X, gives a metric of the contribution of the individual multivariate elements to the accuracy of the solution. Every element of this error criterion, $\epsilon = (\epsilon_1, \cdots, \epsilon_d)$ is

$$\epsilon_l = \sum_{i=1}^{N} \|\psi_l(y_i) - U_{d_l} \psi_l(x_i)\|_2 \tag{8}$$

where U_{d_l} is the l-th row of the Koopman operator matrix.

Recall that for the case of orthogonal polynomials, the way to recover the state is by the inverse of n injective multivariate elements, if these elements are the ones with index one in each of the state variables, the recovery is a linear vector-valued function of the observed values.

[1] The quantity $\|\cdot\|_q$ is not a norm because it does not satisfy the triangle inequality.

Therefore, the threshold $\bar{\epsilon}$ for the elimination of elements is the maximum of the errors with an index equal to one,

$$\bar{\epsilon} = \max(\epsilon \colon \|\alpha_l\|_1 = 1)$$

and the multivariate polynomial elements that stay as a par of the reduced vector-valued function of observables $\Psi_R(x)$ are

$$\Psi_R(x) = \{\Psi(x) \colon \epsilon \leq \bar{\epsilon}\}. \tag{9}$$

The application of the p-q-quasi norms and polynomial error reductions allow for the approximation of the Koopman operator of the underlying systems with significant fewer basis elements than the current methods and increases the accuracy of the operator. This increase in accuracy allows for the training of the operator with less training trajectories.

4.2 Methodology and Accuracy

The proposed method for obtaining the low-order polynomials based on the p-q-quasi norms and the error of the multivariate elements is by applying a greedy approach for the calculation of a suboptimal p-q parameterization, and a second calculation of the operator based on the elimination of the multivariate elements.

The greedy approach to find the suboptimal approximation of the Koopman operator relies on the error from the comparison on the test set of orbits, and the predicted orbits of the operator. Consider (5), where the k-th application of the operator U_d gives the evolution of the observables $\Psi(x)$ up to that time. From this evolved state, consider a matrix $B \in \mathbb{R}^{d\times n}$ with a unitary value per column in the position where an element of the vector-valued function has an index equal to on, i.e., every column of matrix is defined as

$$B_l = \{e_l \colon \|\alpha_l\|_1 = 1\},$$

From this definition and defining the inverse of the n injective function of degree one as $\Psi^{-1}(x)$, the evolution of the states is

$$\hat{x}(k) = \Psi^{-1}\left(B^\top U_d^k \Psi(x(0))\right).$$

The empirical error criterion for a number N_s of test trajectories and a length k_j for every one of these trajectories is

$$\epsilon_T = \frac{1}{\sum_{j=1}^{N_s} k_j}\left(\sum_{j=1}^{N_s}\sum_{k=1}^{k_j}\frac{|x_j(k) - \hat{x}_j(k)|}{|x_j(k)|}\right). \tag{10}$$

The first step of the process is to select a set of q norms and a set of maximum multivariate order p. For each p, sweep along the parameter q calculating the error (10) until there is no improvement in two consecutive calculations, then select the suboptimal parameters from the approximation that gives the least error. The second step is to compute the error (8) for the individual multi-variate elements and select the elements that will stay on the polynomial basis according to (9). This procedure gives the suboptimal low-order/dimension polynomial basis.

5. RESULTS

This paper proposes a method to get a low-order/dimension polynomial basis for the approximation of the Koopman operator via the EDMD algorithm. The advantages of this method are that it gets a linearization in the large, a linear system of expanded dimension that captures the dynamics of the state space from which there is available data.

Table 1. System and simulation parameter

a	42.14	K_{s_1}	7.1
c	116.15	K_{s_2}	9.28
d	268	K_{i_2}	256
μ_{m_1}	1.2	ξ_{in_1}	40
mu_{m_2}	0.74	ξ_{in_2}	175
Δt	0.1	D	0.45
Polynomial		Laguerre	
p sweep		5, 6, 7, 8	
q sweep		0.7, 0.8, 0.9, 1.0 ∞	

For the particular case of the AD process with six equilibrium points, eight trajectories evenly distributed in the state space represent the data set. From this dataset, six are for training, giving a training set of 3306 data points, and two are for testing, with a size of 1930 data points. Table 1 describes the parameters for the simulation and the approximation of the Koopman operator. Figure 2 shows the results of the proposed method applied to the AD process. It shows that the empirical error (10) decreases two orders of magnitude from a full index basis to the suboptimal parameterization where $p = 7$ and $q = 1$. This reduction scheme lowers the maximum order of the polynomial basis from 14 to 7 (50%) and reduces the dimension from 64 to 36 (43%). The second reduction scheme takes

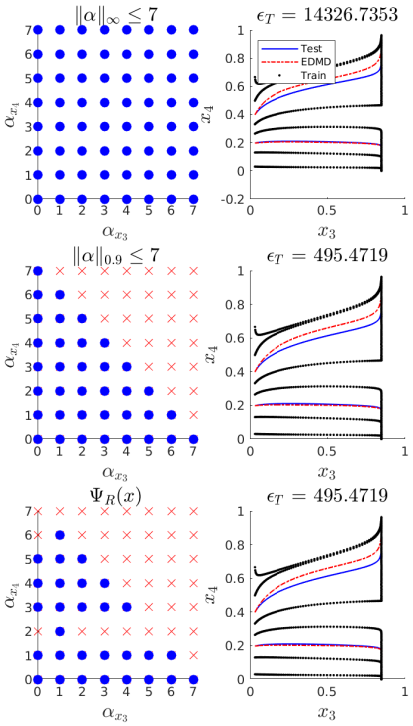

Fig. 2. Indices α for the full polynomial representation, the p-q-quasi norm reduction, and the $\psi(x)$ error reduction with the corresponding orbits for the train, test, and operator orbits.

the suboptimal parameterization and calculates the error (8) per polynomial element and gives a set of observables that remain in the basis (9). The advantage of this process is that the reduction of the dimension does not affect the empirical error.

The importance of these results is twofold. First, the number of initial conditions, and thus the number of data points, to accurately approximate the operator for this unforced representation of the AD process is lower than the traditional methods for approximating the Koopman operator. The conventional methods require hundreds of initial conditions and several thousands of data points to approximate systems of two state variables. Second, the dimension of the polynomial basis dictates the feasibility of the computational implementation of the algorithm. The evaluation of the polynomial basis on every sample and the inverse of matrix G in (4) can hinder the process. Kernel-based methods (Williams et al., 2016b) are useful in these particular cases, but they pose problems when recovering the state and give no information about the stability of fixed points (Garcia-Tenorio et al., 2019)

6. CONCLUSIONS

This paper explores the approximation of expanded linear models of the nonlinear AD process. These linear models have the necessary accuracy in predicting the behavior of the states while keeping the operato'r order and dimension low.

Although the proposed method relies on the measurements of the species concentrations, which is not feasible in practice, it does show that it is feasible to get accurate models for this particular case, when there is a small number of available orbits.

REFERENCES

Bastin, G. and Dochain, D. (1990). *On-line Estimation and Adaptive Control of Bioreactors*. Elsevier, Amsterdam.

Bastin, G. and Van Impe, J. (1995). Nonlinear and Adaptive Control in Biotechnology: A Tutorial. *European Journal of Control*, 1(1), 37–53.

Batstone, D., Keller, J., Angelidaki, I., Kalyuzhnyi, S., Pavlostathis, S., Rozzi, A., Sanders, W., Siegrist, H., and Vavilin, V. (2002). The IWA Anaerobic Digestion Model No 1 (ADM1). *Water Science and Technology*, 45(10), 65–73.

Bernard, O., Hadj-Sadok, Z., Dochain, D., Genovesi, A., and Steyer, J.P. (2001). Dynamical model development and parameter identification for an anaerobic wastewater treatment process. *Biotechnology and Bioengineering*, 75(4), 424–438.

Brunton, S.L., Proctor, J.L., and Kutz, J.N. (2016). Discovering governing equations from data by sparse identification of nonlinear dynamical systems. *Proceedings of the National Academy of Sciences*, 113(15), 3932–3937.

Budišić, M., Mohr, R., and Mezić, I. (2012). Applied koopmanism. *Chaos: An Interdisciplinary Journal of Nonlinear Science*, 22(4), 47510.

Dochain, D., Perrier, M., and Pauss, A. (1991). Adaptive Control of the Hydrogen Concentration in Anaerobic Digestion. *Industrial and Engineering Chemistry Research*, 30(1), 129–136.

Donoso-Bravo, A., Mailier, J., Martin, C., Rodríguez, J., Aceves-Lara, C.A., and Wouwer, A.V. (2011). Model selection, identification and validation in anaerobic digestion: A review.

Garcia-Tenorio, C., Tellez-Castro, D., Mojica-Nava, E., and Wouwer, A.V. (2019). Analysis of a Class of Hyperbolic Systems via Data-Driven Koopman Operator. In *International Conference on System Theory, Control and Computing (ICSTCC)*, 566–571.

Giovannini, G., Sbarciog, M., Steyer, J.p., Chamy, R., and Vande, A. (2018). On the derivation of a simple dynamic model of anaerobic digestion including the evolution of hydrogen *. *Water Research*, 134, 209–225.

Koekoek, R., Lesky, P.A., and Swarttouw, R.F. (2010). *Hypergeometric Orthogonal Polynomials and Their q-Analogues*. Springer Monographs in Mathematics. Springer Berlin Heidelberg, Berlin, Heidelberg.

Konakli, K. and Sudret, B. (2016). Polynomial meta-models with canonical low-rank approximations: Numerical insights and comparison to sparse polynomial chaos expansions. *Journal of Computational Physics*, 321, 1144–1169.

Koopman, B.O. (1931). Hamiltonian Systems and Transformation in Hilbert Space. *Proceedings of the National Academy of Sciences*, 17(5), 315–318.

Korda, M. and Mezić, I. (2018). On Convergence of Extended Dynamic Mode Decomposition to the Koopman Operator. *Journal of Nonlinear Science*, 28(2), 687–710.

Mailleret, L., Bernard, O., and Steyer, J.P. (2003). Robust regulation of anaerobic digestion processes. *Water Science and Technology*, 48(6), 87–94.

Sbarciog, M., Loccufier, M., and Noldus, E. (2010). Determination of appropriate operating strategies for anaerobic digestion systems. *Biochemical Engineering Journal*, 51(3), 180–188.

Schmid, P.J. (2010). Dynamic mode decomposition of numerical and experimental data. *Journal of Fluid Mechanics*, 656, 5–28.

Williams, M.O., Hemati, M.S., Dawson, S.T.M., Kevrekidis, I.G., and Rowley, C.W. (2016a). Extending Data-Driven Koopman Analysis to Actuated Systems. *IFAC-PapersOnLine*, 49(18), 704–709.

Williams, M.O., Rowley, C.W., and Kevrekidis, I.G. (2016b). A kernel-based method for data-driven koopman spectral analysis. *Journal of Computational Dynamics*, 2(2), 247–265.

Yu, L. and Wensel, P.C. (2013). Mathematical Modeling in Anaerobic Digestion (AD). *Journal of Bioremediation & Biodegradation*, s4.