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

Analysis of the ROA of an anaerobic digestion process via data-driven Koopman operator

<https://doi.org/10.1515/nleng-2021-0009>

Received Aug 20, 2020; accepted Mar 23, 2021.

Abstract: Nonlinear biochemical systems such as the anaerobic digestion process experience the problem of the multi-stability phenomena, and thus, the dynamic spectrum of the system has several undesired equilibrium states. As a result, the selection of initial conditions and operating parameters to avoid such states is of importance. In this work, we present a data-driven approach, which relies on the generation of several system trajectories of the anaerobic digestion system and the construction of a data-driven Koopman operator to give a concise criterion for the classification of arbitrary initial conditions in the state space. Unlike other approximation methods, the criterion does not rely on difficult geometrical analysis of the identified boundaries to produce the classification.

Keywords: Anaerobic Digestion, Extended Dynamic Mode Decomposition, Koopman Operator, Region of Attraction

1 Introduction

The biological process through which multiple organisms break down organic matter in the absence of oxygen is called the Anaerobic Digestion process. What motivates the use of the anaerobic digestion 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 production of methane

gas [1]. However, the complexity of the process related to the several inhibition phenomena it difficult to operate because knowledge and expertise are necessary to maintain the stability of the process in the desired working condition. Varying operating variables such as the organic load at the input feed, the dilution rate, or an accumulation of intermediate compounds can destabilize the process leading it to a washout condition [2, 3].

This paper analyzes the inherent difficulties of the process, such as nonlinearities and the multi-stability phenomena with a data-driven linear operator framework. This analysis allows the identification of essential conditions in the operation of the reactor for control purposes. The analysis of multi-stability phenomena present in some nonlinear systems is closely related to the concept of attraction or stability regions. This subject is as important as the notion of stability because the stable operating points of a nonlinear system are rarely globally stable [4]. For the anaerobic digestion process, it is essential to know the attraction region of the desired operating point where the biological species interact and coexist, and the attraction region of the different washout points where bacterial life has disappeared.

There are some traditional methods to approximate the region of attraction (ROA) of nonlinear systems with multiple asymptotically stable equilibrium points. First, Lyapunov based methods approximate the ROA by computing the level sets of the Lyapunov function [5]. Additionally, model-based techniques approximate the ROA by the backward integration of the system from the vicinity of saddle equilibrium points in its boundary. Finally, energy function based techniques approximate the boundary of the ROA by using the level sets of the local energy function (which are constant along the trajectories) [4].

All of these techniques have their inherent difficulties, Lyapunov and energy based techniques require the ability to find complex functions of the state variables that satisfy certain conditions with respect to the nonlinear state space representation. Finding these functions becomes particularly difficult when the dimension of the state space gets higher. The backward integration of the system generates a set of points that represent the stable manifold of saddle points that give the boundary of the ROA. These

***Corresponding Author: Camilo Garcia-Tenorio**, Department of Mechanical and Mechatronic Engineering, Universidad Nacional de Colombia, Bogota, Colombia & Control Systems, Estimation, Control and Optimization (SECO) Lab, Université de Mons, Mons, Belgium, E-mail: cagarciate@unal.edu.co

Eduardo Mojica-Nava, Department of Electrical and Electronic Engineering, Universidad Nacional de Colombia, Bogota, Colombia, E-mail: eamojican@unal.edu.co

Mihaela Sbarciog, Chemical and Biochemical Process Technology & Control, KU Leuven, Gent, Belgium, E-mail: mihaela@sbarciog.be

Alain Vande Wouwer, Systems, Estimation, Control and Optimization (SECO) Lab, Université de Mons, Mons, Belgium, E-mail: alain.vandewouwer@umons.ac.be

points are a set of hyperplanes that ultimately represent the division of the ROA for the different asymptotically stable points. However, the analysis relies on the ability to find conditions according to the geometry of these sets of points to perform an accurate classification of the initial conditions. Similarly to the Lyapunov functions, the complexity of this problem grows with the dimension of the state space.

The first step to determine the proper operation of a reactor and the ROA for feedback control purposes is the development of a dynamic model of the system. There are several such models in the literature, varying from very detailed models with various bacterial populations and substrates [6], to mass-balance systems of two populations and two substrates [7]. These representations give a proper description of the dynamics, even though the task of modeling, parameterizing or calibrating the system by numerical methods is difficult [8]. For the particular case of the AD (see [9] and the references therein), the task depends on the measurements from a live digester or the simulated dynamics of detailed models such as the International Water Association *anaerobic digestion model number one* (ADM1) [6].

In this paper we propose a new approach based on the integration of such numerical representations from different initial conditions to estimate the discrete-time approximation of the Koopman operator. The approximation comes from the extended dynamic mode decomposition (EDMD) [10, 11], where the available spectrum of the approximated operator gives a concise criterion that does not rely on the ability to find or estimate Lyapunov or energy functions, or the geometrical analysis of identified hyperplanes in the state space. The approach relies on the fact that the EDMD algorithm gives a series of eigenfunctions whose associated eigenvalue is unitary. These eigenfunctions are invariant along the trajectories of the system and as a consequence have the ability to capture the stable manifold of saddle equilibrium points that give the boundary of the asymptotically stable points ROA. To achieve this objective in a data-driven way, the proposed algorithm is able to first approximate the discrete-time Koopman operator with the least amount of data. Regarding this approximation, the contribution of this paper is in the description of the p - q -quasi norm, and polynomial accuracy methods that allow the reduction of the necessary basis of orthogonal polynomials for the accurate approximation. Next, it is necessary to approximate the location of the system fixed points. For this purpose, this paper describes how the linear predictors that come from the EDMD algorithm are suitable for the formulation of a multi-start optimization algorithm that accurately locates these points. Addition-

ally, this paper describes the methods to determine their stability. As the EDMD algorithm also gives a non-linear model of the system, the linearization and local analysis of this model gives the fixed points' stability: asymptotically stable, unstable or saddle. Coupled with the identification of the saddle points, and given that there are several of these for the proposed parametrization of the system, this paper shows how to determine which saddle points are in the boundary of the ROA for their subsequent analysis. Finally, the main contribution of this paper is to show how to set a classification criterion for an arbitrary initial condition of the system based on the eigenfunction of the Koopman operator. This criterion determines to which attractor (asymptotically stable point) will an arbitrary initial condition converge with a simple algebraic comparison. It is important to stress that the only use of the differential equation model is in the generation of the sample trajectories that provide the data to the algorithm, and although there is a linearization of the system that the EDMD algorithm provides, this model is still a consequence of the data-driven method.

Although the proposed algorithm needs a sufficient set of data that comes from the numerical integration of a dynamical system representation of the anaerobic digestion process, the fact that it does not subsequently rely on the manipulation of the differential equation model is an advantage over traditional techniques. Indeed, the algorithm can be extended to operate on real measured data from a digester bypassing the difficulties of identifying and parameterizing a dynamical system with the same measurements.

The rest of the paper is organized as follows: in Section 2 we present a detailed description of the dynamics under analysis, and the inherent difficulties that the proposed methodology overcomes over the traditional techniques. Section 3 presents the mathematical tools that we need for the development of the algorithm, including a detailed description of the concept of regions of attraction, the Koopman operator, and the method to get its discrete-time approximation via the EDMD algorithm. In this section we also include some of our previous results for the connection between the eigenfunctions with the approximation of the ROA, and subsequently in Section 4 we present some improvements on the EDMD algorithm for reducing the approximation of the discrete-time Koopman operator and the accuracy of the calculation. Furthermore, in Section 5 we present the results of the method and a detailed description of the necessary steps to perform the analysis. Finally, in Section 6 we present some analysis and conclusions.

Notation

$\mathbb{C}$ denotes the field of complex numbers. $\mathbb{R}$ and $\mathbb{R}_+$ denote the field of real and nonnegative real numbers, respectively. For any matrix $A \in \mathbb{R}^{n \times n}$, A^* denotes complex conjugate transpose, A^+ denotes its pseudoinverse, and $\|x\|$ represents the Euclidean norm. 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 Hadamard product. 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 Problem statement

The model of the anaerobic digestion system under study considers two reactions: acidogenesis and methanogenesis. The following reaction network describes the biological transformations:

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 continuously stirred tank reactor, the following differential equations describe the system dynamics of the reaction network (1):

$$\begin{aligned} \dot{\xi}_1 &= u(\xi_{\text{in}_1} - \xi_1) - ar_1(\xi), \\ \dot{\xi}_2 &= u(\xi_{\text{in}_2} - \xi_2) + cr_1(\xi) - dr_2(\xi), \\ \dot{\xi}_3 &= -u\xi_3 + r_1(\xi), \\ \dot{\xi}_4 &= -u\xi_4 + r_2(\xi), \end{aligned} \quad (2)$$

with a methane outflow rate

$$Q(\xi) = qr_2(\xi), \quad (3)$$

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 dilution rate, $\xi_{\text{in}_1}, \xi_{\text{in}_2} \in \mathbb{R}_+$ are the concentrations of organic substrate and volatile fatty acids in the influent,

$a, b, c \in \mathbb{R}_+$ are the stoichiometric coefficients, $q > 0$ is the yield for the methane production, and $r_1(\xi), r_2(\xi)$ are the reaction rates defined as

$$r_1(\xi) = f_1(\xi_1)\xi_3, \quad (4)$$

$$r_2(\xi) = f_2(\xi_2)\xi_4, \quad (5)$$

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

$$\begin{aligned} f_1(\xi_1) &= \mu_{m_1} \frac{\xi_1}{K_{s_1} + \xi_1}, \\ f_2(\xi_2) &= \mu_{m_2} \frac{\xi_2}{K_{s_2} + \xi_2 + \frac{\xi_2^2}{K_{i_2}}}. \end{aligned} \quad (6)$$

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 $x_b = \xi_b - C_b C_a^{-1} \xi_a$ [8, 12] as

$$\begin{aligned} \dot{x}_a &= u(w_a - x_a) + C_a \rho(x), \\ \dot{x}_b &= u(w_b - x_b), \end{aligned} \quad (7)$$

where $x_a \triangleq [x_3 \ x_4]^T$ 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 a linear combination of the four state variables, $C_a = I_2$, $\rho(x) \triangleq [\rho_1(x) \ \rho_2(x)]$ are the reaction rates, $w_a \triangleq [w_3 \ w_4]^T = [0 \ 0]^T$ are the concentration of species in the input flow, $w_b \triangleq [w_1 \ w_2]^T = [\xi_{\text{in}_1} \ \xi_{\text{in}_2}]$ 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

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

Unlike the four-state representation (2), the canonical state representation is a two-state variable nonlinear differential equation system in the bacterial populations, and a two-state variable linear differential equation. This linear part is assumed to be fast in comparison with the nonlinear part and therefore its dynamics can be neglected for the analysis of the behavior of the system in the species concentration. As a consequence the problem reduces to a two-state system for which the analysis is easier and can be illustrated in 2D figures of the bacterial concentrations.

Table 1: Model parameters

Parameter	Value	Units
a	42.14	
K_{s1}	7.1	g/l
c	116.15	mmol/g
K_{s2}	9.28	mmol/g
d	268	mmol/g
K_{i2}	256	mmol/g
μ_{m1}	1.2	day ⁻¹
ξ_{in1}	[60 – 90]	g/l
μ_{m2}	0.74	day ⁻¹
ξ_{in2}	[160 – 245]	mmol/l
q	453	
u	[0.42 – 0.5]	day ⁻¹

Depending on the magnitude of the dilution rate and substrate concentration at the inflow, the system may possess up to six equilibrium points. In [2] and [13], the authors present the regions of the input space for which various numbers of equilibria occur, and their stability. This paper will consider the multi-stability case, where there are two asymptotically stable equilibrium points, one representing the desired working point of the system, where the acidogenic and methanogenic bacteria coexist, and the other, the acidification point. The objective is to find the attraction region of the coexistence point.

Consider the benchmark presented in [14] with system parameters as in Table 1, where the authors present the stability boundaries for the problem using backward integration. For the analysis, the value of the dilution rate is fixed at $u = 0.45$ [day⁻¹], and the concentration of substrates at the input are set in a grid, where $40 \leq w_1 \leq 70$ [g/l] and $170 \leq w_2 \leq 225$ [mmol/l]. This paper will not consider a unique dilution rate, instead, it considers a three dimensional grid of dilution rate, and input concentrations to perform the analysis.

Figure 1 depicts the phase plane orbits of bacteria population for different values of the inlet for the dilution rate and substrate concentration. The system has six equilibrium points; two asymptotically stable, the working point x_A^* in which the species coexist and the acidification point x_B^* of methanogenic wash-out. Additionally, the unstable saddle points are: the point that separates the two asymptotically stable x_C^* , the two acidogenic wash-off points x_D^* , x_E^* , and the total wash-out of the system x_F^* . Each of these points is of *type-k*, where k is the number of unstable modes of the equilibrium point.

The objective is to get the separation of the state space for the different parameterizations. This separation specifies if an initial condition converges to either of the asymptotically stable equilibrium points.

2.1 Traditional methods

Traditionally, the ROA in the multi-stability problem comes from the maximal Lyapunov set. This is a domain that contains the equilibrium point where the local Lyapunov function is positive definite and its derivative along the trajectories is negative, i.e., the estimate of the ROA is the largest level set of the Lyapunov function where the aforementioned conditions hold [5], where the constant value for this largest level set is often called the critical level value.

Even though there are several optimization methods to find local Lyapunov functions from linear programming or linear matrix inequalities [15], this is not an easy task, especially when the dimensionality of the system grows. Additionally, the level sets of the Lyapunov function are a conservative approach, given that the Lyapunov function is not unique, and it is a greater challenge to find a function whose level sets maximize the estimate of the ROA.

In addition to the Lyapunov methods, energy function methods give a less conservative approximation of the ROA because the conditions that an energy function needs to satisfy are less restrictive than those of Lyapunov functions. For instance, the derivatives of an energy function along the trajectories of the system are non-positive (in contrast with the positive definiteness of Lyapunov functions, and the negative definiteness of their derivatives). The approximation of the ROA comes from the same maximal level set where its critical level value corresponds to the evaluation of the closest saddle point in the boundary with the energy function.

The feasibility of getting a solution with either of these methods is questionable, starting from the difficulties in getting a function that satisfies the conditions, coupled with the difficulties of calculating level sets. In addition, those level sets are points of a $n - 1$ dimensional hyperplane of the state space, where the classification of an arbitrary initial condition on higher than three dimensional systems is not trivial. This classification depends on the geometry of the stable manifold, and the ability to interpolate between the identified points to perform the comparison.

Alternatively, the backward integration method, for approximating the boundary of the ROA proceeds as follows:

- Locate all the equilibrium points of the system and determine their local stability by the Jacobian matrix.
- Find the stable eigenvectors of the saddle points.
- Find the intersection of these eigenvectors with an ϵ -ball of the saddle point, and integrate numerically to

Figure 1: Phase plane with three different realizations of the anaerobic digestion process.

check which of the asymptotically stable points ROA is under analysis.

- Finally, perform the backward integration, or a trajectory reversing algorithm from several initial conditions close enough to the intersection point to get the boundary.

In summary, the method gives a set of trajectories that converge to the saddle point in the boundary. Although this method can yield an accurate approximation of the boundary, it is computationally intensive, and has the same geometry and interpolation problems of the Lyapunov and energy function methods. For a detailed approximation of the ROA in the problem at hand using backward integration we refer to [14].

Within the theoretical framework under consideration, that is, the use of an approximation of the Koopman operator to analyze the system, there are some recent studies [16–19] that lead to our theoretical framework and algorithm. Most notably, in [16] the authors propose a visual method where the time average of an observable (i.e., an arbitrary function of the state space) approximates the eigenfunction of the Koopman operator whose eigenvalue is approximately unitary. Although this method provides visual information in two-dimensional systems or slices of three dimensional ones, the method does not give a criterion to determine the convergence of an arbitrary point, and is not feasible to analyze higher dimensional systems. Another notable approach is [18], where the authors calculate isostables of the system. An isostable of a stable equilibrium point is a set of points, or initial conditions that converge synchronously to the attractor, that is, they simultaneously intersect subsequent isostables along the

trajectory to the stable point. The definition of an isostable comes from the magnitude of the Koopman operator slowest eigenfunction, whose level sets give the isostables. Similar to the Lyapunov and energy function based methods, this approach requires the calculation of the level sets of a particular function, which yields the same problems related to those methods regarding the classification of an arbitrary initial condition in higher dimensional systems.

3 Materials and methods

3.1 Stability boundary

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 state space, $k \in \mathbb{Z}_0^+$ is the discrete-time, and $T: \mathcal{M} \rightarrow \mathcal{M}$ is the differentiable vector-valued evolution map, i.e.,

$$x(k+1) = T(x(k)), \quad x_0 = x(0). \quad (8)$$

The solution to (8) is the successive application of T from an initial condition $x_0 \in \mathcal{M}$ at $k = 0$, i.e., $x_k = T^k(x_0) \in \mathcal{M}$, which is an infinite sequence called a trajectory of the system. Suppose $x^* \in \mathcal{M}$ is a fixed point of (8); i.e., $T^k(x^*) = x^*$. The linearization principle defines the local stability of hyperbolic fixed points, i.e., points that satisfy the Hartman-Grobman theorem¹ [5, 20]. This prin-

¹ In general, if the modulus of the eigenvalues of a discrete-time fixed point is not strictly equal to one, the fixed point satisfies the Hartman-Grobman theorem.

ciple states that a fixed point x_s^* is asymptotically stable if the modulus of all the eigenvalues of the Jacobian matrix evaluated at the fixed point are less than one, and unstable otherwise with an index equal to the number of eigenvalues with modulus greater than one.

Eigenvalues with modulus less than one are stable eigenvalues with stable eigenvectors that span the stable eigenspace of the fixed point, and conversely for the eigenvalues with modulus greater than one. As the Hartman-Grobman theorem establishes a one-to-one correspondence between the nonlinear system and its linearization, locally, the stable and unstable eigenspaces are tangent to the stable and unstable manifolds of the hyperbolic fixed point. The definitions of these manifolds are

$$W^s(x^*) = \left\{ x \in \mathbb{R}^n : \lim_{k \rightarrow \infty} T^k(x) = x^* \right\}, \quad (9)$$

$$W^u(x^*) = \left\{ x \in \mathbb{R}^n : \lim_{k \rightarrow -\infty} T^k(x) = x^* \right\}, \quad (10)$$

for the stable and unstable manifold respectively, assuming there is an inverse for T^k . When the index of unstable fixed points is equal or greater than one, and less than n , the fixed point is called a saddle, denoted by $\hat{x}^*$.

With these definitions, the theorem characterizing the ROA for asymptotically stable fixed points, defined as $R_A(x_s^*)$ with a stability boundary $\partial R_A(x_s^*)$, is taken from [4, Th. 9-(10,11)], where the system satisfies the following assumptions:

- A1 All the fixed points on the stability boundary are saddle points.
- A2 The stable and unstable manifolds of the fixed points on the stability boundary satisfy the transversality condition².
- A3 Every trajectory on the stability boundary approaches one of the saddle points as $k \rightarrow \infty$.

Theorem 1. Consider an autonomous nonlinear discrete-time dynamical system (8), satisfying assumptions (A1-A3). Let $\{\hat{x}_i^*\}_{i=1}^P$ be the P saddle points on the stability boundary of an asymptotically stable fixed point. Then,

- a $\hat{x}_i^* \in \partial R_A(x_s)$ if and only if $W^u(\hat{x}_i^*) \cap R_A(x_s^*) \neq \emptyset$
- b $\partial R_A(x_s^*) = \cup W^s(\hat{x}_i^*)$.

Hence, the stability boundary is the union of the stable manifolds of the hyperbolic equilibrium points on the stability boundary.

3.2 The Koopman operator

The Koopman operator framework provides a set of eigenfunctions that are observables or measures of a particular dynamical system, whose evolution relates to that of the states of the system. This approach allows for an analysis of the eigenfunctions rather than the analysis of the differential equation [21]. The trade-off with this approach concerns linearity and dimensionality, the finite-dimensional non-linear systems being described by an infinite-dimensional linear operator. The linearity of the operator allows having a spectral decomposition that contains information about the underlying dynamical system.

Consider the autonomous nonlinear discrete-time dynamical system (8). The operator U^k , so-called Koopman operator defines the evolution of observable functions $f(x) \in \mathcal{F}$ with $f(x) : \mathcal{M} \rightarrow \mathbb{C}$

$$[U^k f](x) = f(T^k(x)). \quad (11)$$

The spectral decomposition yields tuples $\{(\mu_i, \phi_i(x), v_i)\}_{i \in \mathbb{N}}$ of eigenvalues, eigenfunctions and modes that contain information on the underlying dynamical system. The eigenfunctions and eigenvalues satisfy the condition

$$[U^k \phi_i](x) = \mu_i \phi_i(x), \quad (12)$$

meaning that the evolution of a particular eigenfunction is given by its eigenvalue, and the modes serve to recover the values of the original set of observables

$$f(x) = \sum_{i=1}^{\infty} v_i \phi_i(x). \quad (13)$$

A family of polynomial difference equations has a closed analytical Koopman operator representation given by Bunton in [22]. An extension of this example, where the role of the spectral decomposition is clear, is given in [23].

Given that the closed analytical representation of the Koopman operator is not available for arbitrary dynamical systems, there are available algorithms that calculate an approximation based on data from the numerical integration of a system or real measurements such as the EDMD algorithm [10].

² For manifolds A and B in $\mathcal{M}$, the transversality condition is satisfied if the tangent spaces of the intersection between A and B span the tangent space of $\mathcal{M}$.

3.3 Extended dynamic mode decomposition

The objective of the EDMD algorithm [10] is to get a truncated approximation of the discrete-time Koopman operator based on sampled data of the underlying system [24, 25]. Consider the autonomous nonlinear discrete dynamical system (8), 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

$$X = \begin{bmatrix} x_1 & x_2 & \cdots & x_N \end{bmatrix}, \quad Y = \begin{bmatrix} y_1 & y_2 & \cdots & y_N \end{bmatrix}. \quad (14)$$

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 state space as in the DMD algorithm [11]. The “lifting” procedure consists in evaluating the state space of the system on a vector-valued function of observables $\Psi = [\psi_1 \cdots \psi_d]^T : \mathcal{M} \rightarrow \mathbb{C}^{1 \times d}$ [10]. The approximation of the Koopman operator will be of dimension d , and it will be denoted U_d . This operator has to satisfy the condition

$$\Psi(y) = U_d \Psi(x) + r(x), \quad (15)$$

where $r(x) \in \mathcal{F}$ is the residual term to minimize in order to find U_d . This problem is closely related to the least mean squares problem, polynomial chaos expansion [26], statistical estimation [27], and machine learning [28] problems.

A possible solution to find U_d is the ordinary least-squares solution [10], which has a closed analytical form given by

$$U_d \triangleq AG^+, \quad (16)$$

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

$$G = \frac{1}{N} \sum_{i=1}^N \Psi(x_i) \Psi(x_i)^\top, \quad (17)$$

$$A = \frac{1}{N} \sum_{i=1}^N \Psi(x_i) \Psi(y_i)^\top. \quad (18)$$

This solution is suitable when matrix G is not ill-conditioned. Otherwise, solutions based on the Moore-Penrose pseudoinverse or regularized least-squares give a more accurate result, the former being the most accurate, and the latter being computationally efficient and able to provide sparse approximations of U_d . For the regularized least-squares case, the objective is to find a set of d linear fitting functions $\{h_j(x) = U_d \Psi(x)^T\}_{j=1}^d$ of the “lifted”

space, where j indexes the rows of the operator that minimize the error function:

$$E(U_d) = \frac{1}{N} \sum_{i=1}^N L(\psi_j(y_i), h_j(x_i)) + \alpha R(U_d), \quad (19)$$

where L is the loss function that measures the model fit, R is the regularization term, and $\alpha \in \mathbb{R}_+$ is a nonnegative parameter. The choice of L gives different types of solutions to the regularized least-squares problem [29], while different regularization terms R provide various solutions to the over fitting problem that arises from ill-conditioned problems.

The common choices for the loss function are the Ridge regression, absolute value loss, and support vector machines (epsilon-intensive) functions [29]. For the regularization term, the common choices are l_1 , l_2 , and a linear combination of the two (elastic net) norms.

With the solution to (15), and the spectral decomposition M, Ξ, W^* , where $M = \text{diag}(\mu_1, \dots, \mu_d)$ is the diagonal matrix of eigenvalues, $\Xi = [\xi_1, \dots, \xi_d]$ is the matrix of right eigenvectors, and $W^* = \Xi^{-1}$ is the matrix of left eigenvectors, the eigenfunctions of the Koopman operator $\Phi = [\phi_1, \dots, \phi_d]^\top$ [10] are

$$\Phi^\top(x) = \Psi(x)^\top W^*, \quad (20)$$

and (12) gives the evolution of eigenfunctions according to their eigenvalue as

$$\Phi(x(k)) = M^k \Phi(x_0). \quad (21)$$

With the evolution of eigenfunctions, the right eigenvectors or modes serve to recover the original set of observables

$$\Psi(x) = \Xi \Phi(x), \quad (22)$$

and the time evolution of observables according to the spectrum of the Koopman operator [10] is

$$\Psi(T^k(x)) = \Xi M^k \Phi(x). \quad (23)$$

To recover the state, we define a matrix $B \in \mathbb{R}^{(d \times n)}$ with a unitary value per column in the index of an injective observable-function and zeros otherwise. This matrix selects the functions whose inverse gives the value of the state. If we define this set of functions selected by B as $\Psi_B(x)$, then, with (23), the approximation of the state evolution map $T^k(x) = x_k$ from the initial condition x_0 in terms of the approximation of the Koopman decomposition $\{(\mu_i, \phi_i(x), v_i)\}_{i=1}^d$ is

$$\bar{x}(k) = \bar{T}^k(x_0) = \Psi_B^{-1} \left(B^\top \Xi M^k \Phi(x_0) \right). \quad (24)$$

Given that the approximation of the discrete-time Koopman operator comes from a set of orbits of the system, the comparison between these theoretical orbits against the evolution map of the Koopman operator gives a measure for the algorithms accuracy. The empirical error for a number of N_s trajectories, where each has a number of N data points is

$$e = \frac{1}{N_s} \sum_{i=1}^{N_s} \frac{1}{N} \sum_{k=1}^N \frac{|T^k(x_i) - \bar{T}^k(x_i)|}{|T^k(x_i)|}. \quad (25)$$

A low error over a set of test trajectories of the Koopman state evolution map, gives a Koopman operator representation whose eigenfunctions are accurate enough for the subsequent calculations related to them. An error criterion based on the accuracy of eigenfunctions increases the computational effort and does not improve the results of the approximation.

3.4 Spectrum based region of attraction

The results from Theorem 1 state that the boundary of the ROA is the union of the stable manifold of saddle points in the boundary, meaning that it is necessary to find the fixed points, determine their stability, and get the stable manifold of the saddle points.

An equilibrium point of the discrete-time nonlinear system (8) is an invariant subspace that has the property that whenever the state starts in it, it will remain in it for all future time, i.e., $T^t(x^*) = x^*$, where x^* is an equilibrium point of the dynamical system. If the evolution of the state evolution map (24) is invariant then the system is at a fixed point, and, an optimization algorithm from different initial conditions within the state space gives the approximate location of the fixed points [23].

$$\bar{x}^* = \operatorname{argmin}_x \left\| \Psi_B^{-1} \left(B^T \Xi M^k \Phi(x_0) \right) - x \right\|_2^2. \quad (26)$$

Remark. Equation (24) is a nonlinear, discrete-time evolution mapping, and under the definition of fixed point, i.e., $\bar{T}^k(x^*) = x^*$ it is possible to get the approximation of the hyperbolic fixed points by solving

$$\bar{x}^* \subseteq \{x \in \mathbb{R}^n : \bar{T}^k(x) - x = 0\}. \quad (27)$$

In practice, this procedure poses two problems. First, as the nonlinear mapping (24) comes from a set of observables, the dimension and complexity of these functions affect the possibility of getting a solution in polynomial time. In high order polynomial expansions, even if it is possible to find a solution, the computational time is expensively high.

Second, when there is a solution, it is not clear which elements of the subset represent an actual fixed point of the system, i.e., not all the points of the subset are fixed points of the system, while the converse is true. Considering the data driven case, the fixed points of the system are not known a priori and therefore using this definition to approximate the fixed points is not feasible.

Equation (24) is a nonlinear discrete-time state evolution map, and a classical analysis based on the eigenvalues of its Jacobian matrix evaluated at the fixed points provides information on stability.

The method to get the stable manifold of these points with the spectrum of the Koopman operator is to build a set of eigenfunctions with unitary associated eigenvalue [23]. This eigenfunction is

$$\phi_+(x) = \{\phi_1^{k_1}(x)\phi_2^{k_2}(x) : \mu_1^{k_1}\mu_2^{k_2} = 1\}, \quad (28)$$

where $k_1, k_2 \in \mathbb{C}$. This is possible because the product of two arbitrary eigenfunctions is

$$[U_d \bar{\phi}](x) = [U_d(\phi_1^{k_1}\phi_2^{k_2})](x) \quad (29)$$

$$= ([U_d \phi_1](x))^{k_1} ([U_d \phi_2](x))^{k_2} \quad (30)$$

$$= \mu_1^{k_1} \phi_1^{k_1}(x) \mu_2^{k_2} \phi_2^{k_2}(x), \quad (31)$$

meaning that $\bar{\phi}(x)$ is also an eigenfunction of the discrete-time Koopman operator with associated eigenvalue $\bar{\mu} = \mu_1^{k_1}\mu_2^{k_2}$. With these eigenfunctions, it follows that their level set evaluated at the saddle point gives the stable manifold of the point [23], as stated by

Theorem 2. Let $(\mathcal{M}; T(x); k)$ be a dynamical system that accepts an operator representation $(\mathcal{F}_d; U_d; k)$ based on EDMD. If there exists a nontrivial eigenfunction with associated eigenvalue equal to one defined as $\phi_+(x)$, then the stable manifold of a type-one saddle point is a subset of the level set

$$W^s(\hat{x}^*) \subseteq \{x \in \mathbb{R}^n : \phi_+(x) = \phi_+(\hat{x}^*)\}. \quad (32)$$

Remark. The EDMD gives a set of eigenfunctions of the Koopman operator whose associated eigenvalue is one or close to one. There is a particular eigenfunction with unitary associated eigenvalue that maps the state space into a constant value, i.e., $\phi_{\mu=1} = 1$. This is called the trivial eigenfunction and does not give information about the system trajectories.

This level set gives the ROA boundary, and therefore, it is possible to find a classification rule of the different initial conditions in the state space that converge to any of the

asymptotically stable equilibrium points. A geometrical approach to the classification problem is identifying the hyperplane that represents the boundary. This is a common approach when using the backwards integration of the dynamical system to find the set of points that compose the hyperplane. Although this is a possible approach, classifying an initial condition based on the geometry of the manifold is often difficult, especially when the hyperplane describing the boundary is not a function, or the system is of higher than three dimensions.

The approximation of the discrete-time Koopman operator from the EDMD algorithm using orthogonal polynomials as observables gives a set of eigenfunctions whose eigenvalue is unitary or close to unitary. Considering the difficulties of classifying the initial conditions based on the geometry of the stable manifold of saddle points, we propose two different methods based on this set of eigenfunctions to determine the convergence of an arbitrary initial condition.

Definition 3 (Saddle classifier). *Let $\{\phi_+^i: \mu_i = 1\}_{i=1}^\infty$ be a set nontrivial eigenfunctions of a dynamical system $(\mathcal{M}; T(x); k)$ that accepts an approximation of the discrete-time Koopman operator $(\mathcal{F}_d; U_d; k)$ based on EDMD, and let $\hat{x}^*$ be a type-one saddle point in the boundary of the ROA separating the asymptotically stable points x_A^* and x_B^* . The convergence of an arbitrary initial condition x_0 to either asymptotically stable equilibrium points x_A^* or x_B^* is*

$$\begin{aligned} x_A &\subseteq \left\{ x_0 \in \mathbb{R}^n : \Re[\phi_+(x_0)] \geq \Re[\phi_+(\hat{x}^*)] \right\} \\ x_B &\subseteq \left\{ x_0 \in \mathbb{R}^n : \Re[\phi_+(x_0)] < \Re[\phi_+(\hat{x}^*)] \right\} \end{aligned} \quad (33)$$

Remark. *The points that converge to either asymptotically stable fixed point is a subset and not an equality because not all eigenfunctions with associated eigenvalue equal to one capture this phenomena. The particular characteristics that an eigenfunction must have to satisfy this criteria is still an open question.*

This criterion is closely related to the geometric one, especially because there is the need for a clear separatrix given by the real part eigenfunction evaluation in the boundary saddle point. This implies that for every pair of asymptotically stable fixed points to be classified there must be an eigenfunction that separates the two. When the dimension of the problem starts to grow, especially in the amount of asymptotically stable fixed points, the rule given by Definition 3 becomes a combination of the evaluations of the initial condition, compared to the evaluation of the different saddle points in the boundaries.

Because of the invariant nature of the eigenfunction with associated eigenvalue equal to one along the trajectories of the system, we propose a second criterion to classify the initial conditions that is independent on the aforementioned boundary of the ROA given by the level set. Theorem 1 states that the stable manifold of type-one saddle points can be captured with an eigenfunction with associated eigenvalue equal to one. Extrapolating this concept to the asymptotically stable fixed points, it is possible to find a unitary eigenfunction that is invariant along its ROA.

Definition 4 (AS classifier). *Let $\{\phi_+^i: \mu_i = 1\}_{i=1}^\infty$ be a set nontrivial eigenfunctions of a dynamical system $(\mathcal{M}; T(x); k)$ that accepts an approximation of the discrete-time Koopman operator $(\mathcal{F}_d; U_d; k)$ based on EDMD, let $\{x_I^*\}_{I=1}^L$ be a set of L asymptotically stable fixed points in the space state, and define $C_I(\phi): \mathbb{C} \rightarrow \mathbb{C}$ as an arbitrary function of the observation. The set of points x_I that converge from an arbitrary initial condition x_0 to the fixed points x_I^* is*

$$x_I \subseteq \{x \in \mathbb{R}^n : C_I(\phi_+(x_0)) = C_I(\phi_+(x_I^*))\} \quad (34)$$

Remark. *The ideal case is to have a real valued constant eigenfunction where $C_I(\phi(x)) = \phi(x)$. Although this is theoretically possible, i.e., to find an eigenfunction invariant along the ROA of the asymptotically stable fixed point, the necessary conditions or the characteristics for this to happen is still an open question. For $C_I(\phi(x)) = \angle(\psi(x))$, $C_I(\phi(x)) = |\phi(x)|$ and $C_I(\phi(x)) = \Re(\phi(x))$ there is a chance of getting an accurate classification based on the natural eigenfunction from the EDMD approximation or by constructing the eigenfunction. Nevertheless, the current method to find an eigenfunction that accurately classifies according to either of these functions of ϕ is by checking the necessary conditions for the eigenfunction to be accurate under each rule. As well as the constant case, the conditions that guarantee an accurate behavior is still an open question.*

Remark. *Although this eigenfunction is suitable for the classification of all the approximated state space into its respective point of convergence, note that the equality is replaced by a subset because not all eigenfunctions with an associated eigenvalue equal to one satisfy this criterion. To the best of the authors knowledge, the criteria to make it an equality are still unknown.*

Section 5 exemplifies the different results of performing the analysis with the different classification schemes. In general, the available eigenfunctions with a unitary eigenvalue that come directly from the EDMD algorithm or a

constructed one from (28) give an accurate result by using Definition 3. For the classification based on the asymptotically stable equilibrium points from Definition 4, finding an eigenfunction that satisfies this criterion is not trivial.

4 Polynomial base reduction

The choice of orthogonal polynomials for the approximation of the Koopman operator has the advantage that the rank and dimension of the vector-valued function of observables Ψ can be reduced [30, 31], and the state can be recovered by the inverse of injective observable-functions. A sequence of orthogonal polynomials $\{\pi_\alpha(x)\}_{\alpha=0}^p$ where $\pi_\alpha(x)$ is a univariate (i.e., $x \in \mathbb{R}$) polynomial of degree $\alpha \in \mathbb{N}_+$ up to order p . This sequence is 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 $i \neq j$, and satisfies a particular ordinary differential equation. For example, the set of Laguerre Polynomials is defined over the range $[0, \infty]$, satisfies the ordinary differential equation

$$x \frac{d^2 y(x)}{dx^2} + (1 - x) \frac{dy(x)}{dx} + \alpha y(x) = 0. \quad (35)$$

With these univariate polynomials, every element of the vector-valued function of observables is the tensor product of n univariate polynomials, that is,

$$\psi_l(x) = \prod_{j=1}^n \pi_{\alpha_j}(x_j). \quad (36)$$

Consider a full degree and full dimension vector-valued function of observables, the order and dimension for a particular system grows exponentially with the number of state variables n . The maximum degree of the polynomials in $\Psi(x)$ is p^n and the dimension of the whole set is

$$\dim \Psi(x) = \binom{n+p+1}{p+1}. \quad (37)$$

The exponential growth of the order produces an overfitting problem in the discrete approximation of the Koopman operator via the EDMD algorithm, and the growth in the dimension has the course of dimensionality problem, especially in the definition of the discrete Koopman matrix from (16) that needs the inverse of matrix G . To avoid these problems it is necessary to restrict the maximum order of the polynomial basis and its dimension.

4.1 Reduction by p-q-quasi norms

The reduction of the order and dimension of the polynomial basis proposed by the authors [23] is via a p-q-quasi

norm³ reduction first introduced in the solution of polynomial chaos problems [30]. Define the q-quasi norm as,

$$\|\alpha\|_q = \left(\sum_{i=1}^n \alpha_i^q \right)^{\frac{1}{q}}, \quad (38)$$

for a polynomial element $\psi_l(x)$ consider a maximum degree $p \in \mathbb{N}$, a quasi norm $q \in \mathbb{R}_+$ and define the set of indexes α_l as

$$\alpha_l = \{\alpha \in \mathbb{N}^n : \|\alpha\|_q \leq p\}. \quad (39)$$

Given that each element of the polynomial basis is the tensor product of the univariate polynomials in the retained indexes given by (36), this approach reduces the maximum order, and the overall dimension of the vector-valued function of observables.

4.2 Reduction by polynomial accuracy

Recall from Section 3.3 that the discrete-time approximation of the Koopman operator comes from a set of orbits of the system that gives the linear evolution of a set of observables $\Psi(x)$ given in (15). Consider now the evolution of a single observable function of this set, i.e.,

$$\Psi_l(y) = U_{d_l} \Psi(x) + r_l, \quad (40)$$

where U_{d_l} is the l -th row of the Koopman operator matrix, and r_l is the residual term or error for that particular polynomial element. The overall error of every element in the polynomial basis over a set of orbits of length N gives a metric of the contribution of that particular polynomial element into the accuracy of the approximation of the discrete-time Koopman operator. All the metrics for the individual elements serve to define an overall error criterion $\epsilon = (\epsilon_1, \dots, \epsilon_d)$ as

$$\epsilon_l = \sum_{i=1}^N \|\psi_l(y_i) - U_{d_l} \Psi(x_i)\|_2^2, \quad (41)$$

From the definition of this error, there are some elements that will remain in the basis for the approximation of the Koopman operator because their error is below a threshold, and others that will be eliminated.

The definition of the elimination threshold depends on the selection of the injective functions that must be present in the basis to recover the state. Recall from Section 3.3 that to recover the state we need the matrix B that selects the n injective functions that recover the state.

³ The quantity $\|\cdot\|_q$ is not a norm because it does not satisfy the triangle inequality.

The easiest way to select these injective functions is to choose the order-one polynomial elements in the basis, which gives a set of linear injective functions of the observations whose inverse is linear and easy to calculate. This means that every column of the matrix B is

$$B_n = \{e_n : \|\alpha_n\|_1 = 1\}. \quad (42)$$

As a consequence, the error threshold $\bar{\epsilon}$ that guarantees that the order-one polynomial elements remain in the basis is

$$\bar{\epsilon} = \max(\epsilon : \|\alpha_l\|_1 = 1), \quad (43)$$

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

$$\Psi_R(x) = \{\Psi(x) : \epsilon \leq \bar{\epsilon}\}. \quad (44)$$

In contrast to the previous approach, this reduction only affects the overall dimension of the vector-valued function of observables. Although, this reduction by polynomial accuracy requires two iterations of the EDMD algorithm, increasing the computational burden, reducing the dimension of the linear predictors on observables from the discrete-time Koopman operator approximation is beneficial for the subsequent analysis of eigenfunctions and possible control algorithms.

5 Analysis of the anaerobic digestion process

The previous sections present the necessary tools to perform the analysis, a methodology to get an approximation of the discrete-time Koopman operator based on orthogonal polynomials, and a couple of degree and dimension reduction methods. This representation gives the possibility to approximate the location of the fixed points of the system, provides their stability, and use the set of eigenfunctions with an associated eigenvalue equal to one in conjunction with the *type-one* saddle points in the boundary of the ROA to determine the convergence of an arbitrary point in the state space. In summary, the required steps to perform the analysis are:

- 1 From a set of system trajectories, and with the EDMD algorithm of Section 3.3, calculate the discrete-time approximation of the Koopman operator.
- 2 From the discrete-time representation, use (26) to find the fixed points of the system.
- 3 Calculate the Jacobian of (24), and evaluate it on the fixed points to get their local stability according to the

Table 2: Algorithm Parameters.

u	[0.42, 0.44, 0.44, 0.5]		
ξ_{in_1}	[60 67.5 75 82.5 90]		
ξ_{in_2}	[160.0 181.25 202.5 223.75 245.0]		
Polynomials	Hermite		
	Legendre		
	Laguerre		
	Chebyshev first		
	Chebyshev second		
	Jacobi (1, 1)		
	Jacobi (0.5, 1)		
p	[8 9 10 11]		
q	[0.6 0.7 0.8 0.9 1.0 1.1]		
	Min	Max	Average
Train points	2,943	5,237	3,811
Test points	3,253	5,877	4,199
Δt	0.1 [s]		

spectral decomposition of the matrix. It is important to remark that these eigenvalues come from the classical stability analysis of linear systems, and are not related to the eigenvalues from the Koopman operator.

- 4 Analyze the ROA of the system, this includes finding an eigenfunction with associated eigenvalue equal to one for the case where the EDMD algorithm fails to provide an accurate one. This eigenfunction comes from (28).

Given that the anaerobic digestion system in its reduced form (2) is bistable under some operating conditions, where the separatrix between the stability regions is the stable manifold of a unique *type-one* saddle point, we can show the accuracy of the method analyzing this system.

Consider 100 different operating conditions of system (2) in a 3D grid of four different dilution rates u , five different values for the organic substrate ξ_{in_1} and five different values for the volatile fatty acids ξ_{in_2} with values given in Table 2.

5.1 Train and test data

The EDMD algorithm captures the dynamics of the portion of the state space where the data for the calculation lies. For the case of the eigenfunctions that approximate the ROA of the process, it is especially important to have an even distribution of data on the portion of the state space subject to evaluation.

There are some geometrical properties that hold for the reduced Anaerobic Digestion System [2]. Specifically, it is important to note that the reduced system (2) is non-negative by definition; concentrations cannot be negative, and the canonical representation is also compartmental, where all the trajectories lie inside an invariant set S_x given by

$$S_x = \left\{ x \in \mathbb{R}^4 : \begin{array}{l} \xi_1 = x_1 - ax_3 \geq 0 \\ \xi_2 = x_2 + cx_3 \geq 0 \\ \xi_3 = x_3 \geq 0 \\ \xi_4 = x_4 \geq 0 \end{array} \right\}. \quad (45)$$

From the boundary given by this subspace of $\mathbb{R}^4$ we define a set of initial conditions for every one of the cases. Figure 2 depicts the selection of such initial conditions and their respective orbits for an arbitrary realization, specifically, for $u = 0.44$, $\xi_{in_1} = 90$, and $\xi_{in_2} = 245$. The uniform distribution of these initial conditions on the boundary gives enough snapshot data for the approximation and testing of the discrete-time Koopman operator.

Figure 2: Training and testing sets of snapshots for the approximation of the discrete-time Koopman operator. Initial conditions are uniformly distributed in the boundary of S_x

Given that these trajectories come from the numerical integration of (7), an important consideration is the final integration time. The integration algorithm⁴ stops when a specific orbit reaches steady state because repeated data

in the asymptotically stable fixed points affects the condition number of G in (17), and as a consequence it affects the accuracy of the approximation of the Koopman operator from (16). When the integration stops, the final Δt of the integration is inconsistent with the fixed time-step of the integration, therefore, affecting also the accuracy of the approximation. The solution to this problem is to discard the last value of every trajectory.

5.2 Approximating the Koopman operator

The first step to approximate the discrete-time Koopman operator is to define a set of orthogonal polynomials that act as observables for the system. Consider for example a set of Jacobi polynomials $\pi_{\alpha}^{(\eta, \nu)}(x)$ of degree α with parameters η and ν that satisfy the differential equation

$$0 = (1 - x^2) \frac{d^2}{dx^2} y(x) + (\eta - \nu - (\eta + \nu + 2)x) \frac{d}{dx} y(x) + \alpha(\alpha + \eta + \nu + 1)y(x). \quad (46)$$

The solution of this differential equation for a set of non-negative integers $\alpha_1 = \{0, 1, 2, 3\}$, and parameters $\eta = 0.5$ and $\nu = 1$ is summarized in Table 3, where the subscript on α indicates the univariate solutions for the first component of the state vector.

Table 3: Basis for the first state variable x_1 . Solution to the Jacobi ODE with parameters $\eta = 0.5$ and $\nu = 1$.

$J(\alpha_1)$	$\pi_{\alpha_1}^{0.5, 1}(x_1) = y(x_1)$
$J(0_1)$	1
$J(1_1)$	$7x_1/4 - 1/4$
$J(2_1)$	$99x_1^2/32 - 9x_1/16 - 21/32$
$J(3_1)$	$715x_1^3/128 - 143x_1^2/128 - 319x_1/128 + 27/128$

Every element of the polynomial basis comes from (36), which gives a total of 16 polynomial elements for the full basis without truncation that correspond to all the available combinations of the products between two univariate polynomials, i.e., the whole basis is

$$\Psi(x) = \begin{bmatrix} J(0_1)J(0_2) \\ J(1_1)J(0_2) \\ J(2_1)J(0_2) \\ \vdots \\ J(0_1)J(1_2) \\ J(1_1)J(1_2) \\ \vdots \\ J(3_1)J(3_2) \end{bmatrix} = \begin{bmatrix} 1 \\ 7x_1/4 - 1/4 \\ 99x_1^2/32 - 9x_1/16 - 21/32 \\ \vdots \\ 7x_2/4 - 1/4 \\ (7x_1/4 - 1/4)(7x_2/4 - 1/4) \\ \vdots \\ (715x_1^3/128 + \dots + 27/128)(715x_2^3/128 + \dots + 27/128) \end{bmatrix} \quad (47)$$

The first reduction method is to apply a truncation scheme on this basis. From the definition of the q-quasi norm in (38), and the p-q-quasi norm truncation scheme in (39), take for example the values of $p = 3$ and $q = 0.5$ to reduce the degree and dimension of the basis. Figure 3 depicts the result of applying the truncation scheme, where the resulting number of polynomial elements is 8, effectively reducing the maximum degree from 6 to 3 and the dimension by half. The redefined basis for the observables is

$$\Psi(x) = \begin{bmatrix} J(0_1)J(0_2) \\ J(1_1)J(0_2) \\ J(2_1)J(0_2) \\ J(3_1)J(0_2) \\ J(0_1)J(1_2) \\ J(1_1)J(1_2) \\ J(0_1)J(2_2) \\ J(0_1)J(3_2) \end{bmatrix} = \begin{bmatrix} 1 \\ 7x_1/4 - 1/4 \\ 99x_1^2/32 - 9x_1/16 - 21/32 \\ 715x_1^3/128 - 143x_1^2/128 - 319x_1/128 + 27/128 \\ 7x_2/4 - 1/4 \\ (7x_1/4 - 1/4)(7x_2/4 - 1/4) \\ 99x_2^2/32 - 9x_2/16 - 21/32 \\ 715x_2^3/128 - 143x_2^2/128 - 319x_2/128 + 27/128 \end{bmatrix} \quad (48)$$

Figure 3: Full basis vs. truncated basis

With the definition of the train, test and observables set, the discrete-time approximation of the Koopman operator comes from the calculation of matrices A from (18), G from (17), and finally U_d from (16).

Recall that for the state evolution map of the approximation of the Koopman operator (24), and the error threshold (43) for the second reduction method, an elemental matrix B with l rows, and n columns with a single unitary element per column in the index of an injective observable needs to be defined to recover the state and set the error threshold for the elimination of polynomial elements. A simple solution to this problem is to select the univariate polynomial elements of degree one as stated in (42) from the basis. This type of selection also gives the advantage that the inverse of the selected function by B is linear. By applying this criterion, in our application example,

$$B = \begin{bmatrix} 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \end{bmatrix}^T, \quad (49)$$

that selects the second and fifth element of the polynomial basis which are the univariate, order-one, and injective linear functions of the set of observables that define the set $\Psi_B(x) = [\psi_2(x), \psi_5(x)]$, and whose inverse gives the linear function of observations that recover the state as

$$\Psi_B^{-1} = \left[\frac{1}{7}(4x_1 + 1) \quad \frac{1}{7}(4x_2 + 1) \right]. \quad (50)$$

With the definition of B , the approximation is further reduced via the second method, where the threshold for the elimination of polynomial elements from the basis comes from (41), and (43) as

$$\bar{\epsilon} = \max \left(\sum_{i=1}^N \|\psi_2(y_i) - U_{d_2} \Psi(x_i)\|_2^2, \quad (51)$$

$$\sum_{i=1}^N \|\psi_5(y_i) - U_{d_5} \Psi(x_i)\|_2^2 \right), \quad (52)$$

giving a lower degree, and reduced order approximation of the discrete-time Koopman operator that accurately satisfies condition (15) for the set of observables.

With the reduced discrete-time Koopman operator matrix U_d , the set of observables $\Psi(x)$, the set $\Psi_B(x)$, and its corresponding inverse, the linear predictor that approximates the dynamics of the system from an arbitrary initial condition comes from (24). Although it is useful to have this linear approximation of the evolution of the states according to the spectral decomposition of the discrete-time approximation of the Koopman operator, in practice, it is enough to use a linear predictor based on the original set of observables as in (15). Considering again that $y = T^k(x)$,

$$\Psi_B(x_k) = (U_d^k \Psi(x_0))^T B \quad (53)$$

$$y = \Psi_B^{-1}(\Psi_B(x_k)). \quad (54)$$

Equation (53) is linear, and it describes the linear evolution of observations of the state space. Along with matrix B , it only describes the linear evolution of a set of injective functions whose inverse recovers the state. Given that B selects the order-one observables, (54) is also a linear function. The benefit of having the evolution according to the polynomial basis of observables instead of the eigenfunctions of the discrete-time approximation of the Koopman operator is in terms of accuracy. Equation (24) needs the left eigenvectors W^* of matrix U_d to get the eigenfunctions as in (20), and the right eigenvectors Ξ to recover the value of the evolved observations. To have this full representation it is necessary to calculate the inverse of Ξ , adding additional error because of the numerical approximation of the higher order inverse. As a consequence, the set of eigenfunctions also suffer accuracy issues. For eliminating this induced numerical error, the solution is to get matrix

W^* as the right eigenvectors of U_d^T , and apply (20) to get the eigenfunctions.

Using the linear evolution of observables (53) and (54) instead of eigenfunctions, the linear predictor for the case of Jacobi polynomials (48) is

$$\begin{aligned} \Psi_B(x_k) &= (U_d^k \Psi(x_0))^T B \\ &= \begin{bmatrix} U_{d_2} \Psi(x_0) & U_{d_5} \Psi(x_0) \end{bmatrix} \end{aligned} \quad (55)$$

$$y = \begin{bmatrix} \frac{1}{7}(4U_{d_2} \Psi(x_0) + 1) & \frac{1}{7}(4U_{d_5} \Psi(x_0) + 1) \end{bmatrix}. \quad (56)$$

In conclusion, these changes in the computation of the approximation give a linear predictor and a set of eigenfunctions that do not have an additional inverse numerical error.

There are still some open problems to solve in the approximation of the discrete-time Koopman operator. For our particular case, in which we select a type of orthogonal polynomial and then apply a p-q-quasi norm truncation scheme, there is no current solution to determine the type of polynomial, and the values of the p-q parameters that give the best performance. Therefore, we perform the approximation of the discrete-time Koopman operator for different types of polynomials with different values for the p-q parameters and evaluate their performance using the empirical error (25) to select the suboptimal polynomial type and parameter values. Table 2 shows the different type of polynomials, the values for the p-q sweep, and the minimum, maximum, and average number of training and testing points.

Table 4 shows the result of averaging the empirical error given by every one of the polynomials, and the total time required to perform the calculation⁵. Despite the fact that a Legendre polynomial has the lowest average empirical error within the set, there are some cases where the accuracy of the Legendre polynomial basis does not hold well enough to determine an accurate classifier.

Figure 4 shows the empirical error for each of the parameterizations in the grid, for the case, $u = 44$, $\xi_{in_1} = 82.5$ and $\xi_{in_2} = 223.75$, the Legendre polynomial has an error of 1.65, that is high enough for the EDMD algorithm not to give an accurate approximation of the discrete-time Koopman operator. Therefore, the selection of the type of basis for the whole analysis is Jacobi with parameters $\eta = 0.5$, and $\nu = 1$. Note that some of the cases have a larger empirical error in which the approximation is not ideal. The reason to have these inconsistencies in the results will be covered in the next section.

⁵ Intel(R) Core(TM) i7-7567U CPU @ 3.50GHz, 16GB DDR4 RAM @ 2400MHz

Table 4: Grid optimization results for different types of polynomials.

Polynomial	Av. error	Time [m]
Hermite	0.89	31
Legendre	0.64	31
Laguerre	2.75	44
Chebyshev first	0.77	29
Chebyshev second	0.76	30
Jacobi (1, 1)	0.71	31
Jacobi (0.5, 1)	0.67	42

Figure 4: Empirical error for every test of the 3D grid of parameterizations.

5.3 Discrete-time approximation result

Recall from Theorem 1, and Definitions 3 and 4 that it is necessary to have an accurate location and stability of the fixed points of the system. Either to determine the stable manifold of the *type-one* saddle points in the boundary of the ROA for the classification criterion 3, or to apply the classification criterion 4 on the asymptotically stable ones.

For either of the classification criteria to perform correctly it is only necessary to get the location and stability of the working point x_A^* , the acidification point x_B^* , and the point that separates these two x_C^* . Recall from 3.4 that the fixed points are given by the solution of (26), and in practice, the solution to this equation from different start points gives the whole set of fixed points according to the discrete-time approximation of the Koopman operator. The start points for the identification of the fixed points comes from the trajectories of the test set, where the initial

condition of every trajectory is a starting point for the solution. In addition to these points, and for this particular case it was necessary to select a random point along these trajectories because counting only with the initial conditions, there was no available starting point that gave an identification of the acidification wash-out point.

For the stability of the fixed points, consider the linear evolution of observations (56). Although this is a linear function of observations, it is also a nonlinear function of the state that gives an approximation of its nonlinear evolution operator. From traditional control systems theory, the analysis of the eigenvalues from the linearization of the nonlinear evolution mapping evaluated at the fixed points gives their local stability. Therefore, as proposed before by the authors [23],

$$x(k+1) = \left[\frac{\partial \tilde{T}(x)}{\partial x_1} \dots \frac{\partial \tilde{T}(x)}{\partial x_n} \right]_{x^*} x(k) \quad (57)$$

$$= J|_{x^*} x(k) \quad (58)$$

$$= Hx(k), \quad (59)$$

where J is the Jacobian matrix and, H is the Jacobian evaluated at a fixed point. The local stability of a fixed point x^* with respect to the eigenvalues $\{\mu_i\}_{i=1}^n$ from the spectral decomposition of the matrix H are given by,

- if $|\lambda_i| < 1$ for all $i = 1, \dots, n$ then x^* is asymptotically stable.
- if $|\lambda_i| > 1$ for all $i = 1, \dots, n$ then x^* is unstable.
- if $|\lambda_i| > 1$ for some $i = 1, \dots, n_k$ and $|\mu_i| < 1$ for some $i = n_k + 1, \dots, n$, then x^* is also unstable but has modal components that converge to it, making it a saddle point.

The application of (58) on (56), and with $U_d = (U_{i,j})$ the Jacobian matrix $J = (J_{i,j})$ for the nonlinear evolution operator on the states is

$$\begin{aligned} J_{1,1} = & U_{2,2} + \frac{1}{4} U_{2,6} (7x_2 - 1) \\ & + \frac{4}{7} \left(U_{2,3} \left(\frac{99x_1}{16} - \frac{9}{16} \right) \right. \\ & \left. - U_{2,4} \left(-\frac{2145x_1^2}{128} + \frac{143x_1}{64} + \frac{319}{128} \right) \right) \end{aligned} \quad (60)$$

$$\begin{aligned} J_{1,2} = & U_{2,5} + \frac{1}{4} U_{2,6} (7x_1 - 1) \\ & + \frac{4}{7} \left(U_{2,7} \left(\frac{99x_2}{16} - \frac{9}{16} \right) \right. \\ & \left. - U_{2,8} \left(-\frac{2145x_2^2}{128} + \frac{143x_2}{64} + \frac{319}{128} \right) \right) \end{aligned} \quad (61)$$

$$\begin{aligned} J_{2,1} = & U_{5,2} + \frac{1}{4} U_{5,6} (7x_1 - 1) \\ & + \frac{4}{7} \left(U_{5,3} \left(\frac{99x_2}{16} - \frac{9}{16} \right) \right. \end{aligned}$$

Table 5: Anaerobic digestion fixed points for $u = 0.47$, $x_{in_1} = 75$, and $x_{in_2} = 160$.

	Theoretical	Algorithmic	
x_A^*	(1.67, 1.25)	(1.67, 1.25)	Working point
x_B^*	(1.67, 0.00)	(1.67, 0.00)	Acidification
x_C^*	(1.67, 0.86)	(1.67, 0.86)	Separation saddle
x_D^*	(0.00, 0.13)	(0.00, 0.19)	Acidogenic wash-out
x_E^*	(0.00, 0.53)	(0.00, 0.53)	Acidogenic wash-out
x_F^*	(0.00, 0.00)	(0.00, 0.02)	Wash-out

$$-U_{5,4} \left(-\frac{2145x_2^2}{128} + \frac{143x_2}{64} + \frac{319}{128} \right) \quad (62)$$

$$J_{2,2} = U_{5,5} + \frac{1}{4} U_{5,6} (7x_1 - 1) + \frac{4}{7} \left(U_{5,7} \left(\frac{99x_2}{16} - \frac{9}{16} \right) - U_{5,8} \left(-\frac{2145x_2^2}{128} + \frac{143x_2}{64} + \frac{319}{128} \right) \right). \quad (63)$$

Evaluating the identified fixed points in the Jacobian matrix and getting the respective eigenvalues give their local stability. Table 6 shows the result of this process for the same arbitrary case of the anaerobic digestion system. These result comes from a Jacobi polynomial basis with $\eta = 0.5$, $\nu = 1$, $p = 9$, and $q = 0.7$.

Figure 5 shows the theoretical, and approximated fixed points of an arbitrary realization of the anaerobic digestion system from the solution of (5) from 24 initial conditions. These come from the orbits of the train set, one for every starting point of the train set, and one from an arbitrary point along those trajectories. The error of the working point, the separating saddle and the acidification point is low enough for every realization of the system for an accurate analysis of the ROA.

The results of approximating the discrete-time Koopman operator are summarized in Figure 6 that shows in the first row the most accurate realizations of the anaerobic digestion system, and in the second row, the realizations with the highest empirical error. Although the trajectories for the training set of each of the cases come from the same parameterization for the process with a different value for the dilution rate and input concentrations, not all of the orbits for the training set give the same amount of information. The reduced, nonlinear canonical model for the anaerobic digestion system (7) has a movement and bifurcation phenomena for the fixed points depending on the input parameters [2]. For the bi-stability case, the displacement of the separating saddle x_C^* is directly proportional to the values of input concentrations, especially for the volatile fatty acids concentration ξ_{in_2} (see Figure 1). As a consequence, the size of the ROA for the asymptotically stable fixed points changes. The fact that the selection of

initial conditions for all the realizations is uniform along the boundary of the state space, makes the available data for the acidification ROA scarce, and reduces the accuracy of the algorithm. A solution to this problem is to generate the data with an experimental design that gives enough information on both regions. The reason to avoid that approach is because, it defeats the purpose of analyzing the system without prior knowledge of the system, otherwise we would have to know the ROA to find the ROA.

5.4 Approximating the ROA

The spectral decomposition of the discrete-time approximation of the Koopman operator gives a set of eigenvalues with their respective eigenfunctions. In the approximation from a Jacobi-type polynomial basis, and under the selected p-q sweep, each of the system gets a subset of the eigenfunctions with a unitary or near unitary eigenvalue. For the reduced-order anaerobic digestion system, and under the selected set of orbits for the approximation of the Koopman operator, the set of eigenfunctions is rich enough so that it contains at least one unitary eigenfunction that characterizes the stable manifold of the separating saddle point for the classification of an arbitrary initial condition. If that was not the case, the eigenfunction with unitary associated eigenvalue comes from (28). Several tests over the construction of unitary eigenfunctions suggest that selecting the eigenfunctions with a real valued eigenvalue closest to one improves the performance of the algorithm. Also note that (28) has an infinite number of solutions, and as a consequence not all eigenfunctions constructed this way give an accurate result. The selection of the base eigenfunctions and the solutions to the construction to guarantee an accurate approximation for either of the selection criteria is still an open question.

Also, it is important to note that the constant eigenfunction is also part of the unitary eigenfunctions set, i.e., $\phi_+(x) = 1$. In particular, because of the presence of the constant observable when approximating the discrete-time Koopman operator, this trivial eigenfunction is often

Table 6: Anaerobic digestion fixed points for $u = 0.47$, $x_{in_1} = 75$, and $x_{in_2} = 160$.

	$ \lambda_1 $	$ \lambda_2 $	Stability
x_A^*	0.6448	0.8210	AS
x_B^*	0.6444	0.9834	AS
x_C^*	0.6443	1.0212	Saddle
x_D^*	1.0160	1.0596	Unstable
x_E^*	0.9318	1.0597	Saddle
x_F^*	0.9554	1.0589	Saddle

Figure 5: Location of fixed points in the state space from the solution of the minimization problem (26) from different initial conditions.

present in the approximation. As a result, the eigenfunction that gives an accurate classification is a near unitary, or a constructed one.

In our application example, and for the 100 different parameterizations, the approximation of the discrete-time Koopman operator gives a set of eigenfunctions whose unitary or near unitary eigenvalue give an accurate classification by the saddle selection rule (3). This result is depicted in Figure 7, where the associated eigenvalue is real, exactly one, and whose associated eigenfunction is non-trivial. The accuracy of this eigenfunction and the application of the saddle classification rule is the comparison of an additional test set of 900 initial conditions of the state space against the algorithm. For the realization under consideration, the error is 3 misclassified initial conditions out of the 900.

Consider next a realization where the approximation of the discrete-time Koopman operator gives a real valued nontrivial eigenfunction whose associate eigenvalue is near unitary, i.e., $\mu = 1.0078$. Although the accuracy decreases, with the number of misclassified initial conditions being seven out of 900, it is still a good approximation of the stable manifold and the convergence of an arbitrary initial condition in the state space.

For illustration purposes, consider now a case where the associated eigenvalue is not real-valued or unitary, e.g., $\mu = 1.003 + 0.0101i$. The eigenfunction along with the saddle classification algorithm gives an error of 35 misclassified initial conditions out of 900. The result is given in Figure 9 and shows that the eigenfunction is not monotonic with respect to x_4 , and the non-monotonic portion is completely below the threshold given by the real part of the eigenfunction evaluated at the saddle point. If that was not the case the eigenfunction would give a big classification error for the ROA of x_A^* , meaning that it would classify all the initial conditions in the vicinity of x_A^* as convergent into x_B^* .

Recall from Section 2 that the driving forces in the reduced anaerobic digestion process are the dilution rate, which is the controlled input of the process, and the uncontrolled fluctuations of inlet organic matter and volatile fatty acids concentration. These driving forces produce a displacement of the fixed points, and consequently a displacement of the stable manifold of the separating saddle point in the boundary of the ROA [13]. Additionally, consider the canonical system representation as it allows for graphical representations of the bacterial dynamics in a 2D plane. As a consequence the movement of the separating saddle stability boundary can be illustrated in $x_3 - x_4 - x_i$

Figure 6: Training orbits, testing orbits, and fixed points with their respective approximations based on Koopman linear predictor and analysis algorithms.

Figure 7: Theoretical, saddle based classification of initial conditions, and eigenfunction with the corresponding saddle point and stable manifold.

3D plots where $x_i = \{u, \xi_{in_1}, \xi_{in_2}\}$ that sweep over one of the parameters while the other two are constant.

Figure 10 illustrates two different cases for the $x_3 - x_4 - u$ case, varying the dilution rate between $0.42 \leq u \leq 0.54$ as shown in Table 2 and keeping constant values of $\xi_{in_1} = \{90, 82.5\}$, and $\xi_{in_2} = \{160, 224\}$. The plot shows the displacement of the fixed points, especially how the stable manifold and the working point x_A^* move toward each other in the x_4 axis, decreasing the methane production rate according to (3), and reducing the portion of the state space that converges to the working point.

Consider now the $x_3 - x_4 - \xi_{in_1}$ that describes an increase in the organic matter input concentration as $60 \leq \xi_{in_1} \leq 90$ for the values given in Table 2 for two different constant values of dilution rate and volatile fatty acid concentration of $u = \{0.447, 0.473\}$, and $\xi_{in_2} = \{160, 224\}$, using the same fixed values for the input concentration as the previous case. Figure 11 depicts the approximation of the stable manifold of the dividing saddle, and shows that the relation between the organic matter and the dividing saddle is directly proportional in both axis. As a consequence, an increase in the organic matter concentration

Figure 8: Theoretical, saddle based classification of initial conditions, and eigenfunction with the corresponding saddle point and stable manifold.

Figure 9: Theoretical, saddle based classification of initial conditions, and eigenfunction with the corresponding saddle point and stable manifold.

at the input has the potential of increasing the methane production with the drawback of having a reduction of the working point ROA.

The final case considers the same constant values for the organic matter concentration at the input feed of $\xi_{in1} = \{90, 82.5\}$ as in the first case, and evaluates two different values of the dilution rate $u = \{0.447, 0.5\}$ while varying the different values of volatile fatty acids concentration in the input feed given by Table 2, i.e., $160 \leq \xi_{in2} \leq 245$. Figure 12 shows the approximation of the dividing saddle stable manifold that gives the boundary of the ROA for the asymptotically stable fixed points. In this case, the movement of the stable manifold and the working point are on the x_4 axis and is directly proportional to the volatile fatty acids concentration at the input. As a consequence, there is an increase in the methane production and a reduction of the effective state space that converges to the working point.

6 Discussion

In general, not only for the case of the anaerobic digestion system, the analysis of the ROA is static, i.e., it considers a fixed value of the input. For the process at hand these are the input concentrations of substrates and the dilution rate. The rationale behind this approach is the movement and bifurcation phenomena present in the equilibrium points of nonlinear systems. For example, consider the movement of the fixed points with respect to the input concentrations and dilution rate as depicted in Figures 10 to 12. The working point ROA size is inversely proportional to all the variations in the input, while the converse applies for the movement of the working point. Meaning that a transient and steady state optimization of biogas production benefits from a dynamical analysis instead of a static one, which can reduce the computational burden for this type of optimizations, and can improve the current heuristic approach [32].

Figure 10: Intersection of the stable manifold, different realizations with different u , and constant concentrations at the input.

Under the Koopman operator approach and the discrete-time approximation given by the EDMD algorithm, it is possible to formulate an expansion of the state space defined as the product of the original space and the input space, which is equivalent to setting the inputs as state variables without specific dynamics. Using a polynomial expansion on the input space rather than considering the approximation given by Korda [33] as an input affine system gives the possibility to handle it as a closed one, and as a consequence, analyze the movement of the equilibrium points and the stability boundary in a dynamic way.

The approximation of the discrete-time Koopman operator, either in static or dynamic ways via the EDMD algorithm still has some important problems to be resolved. The first one is the amount of necessary data to have an accurate representation. For the anaerobic digestion process at hand, the analysis is possible based on a process model that can be numerically integrated from a set of initial conditions. Even though the reduction methods and the possibility to test with different polynomials give an accurate

representation with a reasonable amount of data, the analysis would be unfeasible if the solely available trajectories were from a real digester.

Additionally, the type of data available also poses a problem for the analysis given that the species concentrations are not commonly available, whereas biogas outflow is. This implies that the algorithm needs to be extended so that the eigenfunctions from an approximation with partial information are still able to capture the dynamic behavior of the system to approximate all the necessary information: fixed points, stability, unitary eigenfunctions, and classification criteria. Furthermore, for the dynamical analysis, it is necessary to get a classification criteria that not only depends on the current state but also depends on the input concentrations and dilution rate of the system.

7 Conclusion

In this paper we replicate the analysis of the region of attraction for the reduced anaerobic digestion process using

Figure 11: Intersection of the stable manifold, different realizations with different u , and constant concentrations at the input.

only data from a set of trajectories of the dynamical system. Specifically, we use the extended dynamic mode decomposition algorithm to approximate a discrete-time Koopman operator. This operator has a spectrum of eigenfunctions where a particular set of these, which have an associated eigenvalue equal to one, are the basis for formulating a clear algebraic criterion to classify an arbitrary point in the state space into either of the asymptotically stable equilibrium points. In contrast to the traditional methods, this approach has the benefit of being strictly data-driven. The analysis provides a clear criterion that does not depend on the specific geometry of a manifold or the ability to calculate level sets and interpolate between data-points for the classification of an arbitrary point in the state space which is the case for any method that relies on the approximation of level sets. As a result, the analysis can be extended to models with more than two reactions and state variables while preserving a clear and simple criterion for the classification.

Given that it is difficult and costly to acquire data from a reactor, this approach also opens up the possibility of

developing an analysis based on the commonly available measurements such as the flow of biogas at the output. To make that possible, it is necessary that the approximation and reduction techniques presented here are extended so that they achieve an accurate approximation of the discrete-time Koopman operator for a reduced set of forced system trajectories.

Funding information: The authors state no funding involved.

Author contributions: All authors have accepted responsibility for the entire content of this manuscript and approved its submission.

Conflict of interest: The authors state no conflict of interest.

Figure 12: Intersection of the stable manifold for different realizations with different x_{in2} , and constant dilution rate, and x_{in1} concentration at the input.

References

- [1] Mailleret L, Bernard O, Steyer JP. Robust regulation of anaerobic digestion processes. *Water Sci Technol.* 2003;48(6):87–94.
- [2] Sbarciog M, Loccufier M, Noldus E. Determination of appropriate operating strategies for anaerobic digestion systems. *Biochem Eng.* 2010;51(3):180–188.
- [3] Alzate-Ibañez AM, Ocampo-Martinez C, Alzate CAC, Montoya VMT. Risk index to monitor an anaerobic digester using a dynamic model based on dilution rate, temperature, and pH. *Nonlinear Eng.* 2020;9(1):35–50.
- [4] Chiang HD, Alberto LFC. *Stability Regions of Nonlinear Dynamical Systems: Theory, Estimation, and Applications.* Cambridge University Press; 2015.
- [5] Khalil HK. *Nonlinear systems.* 3rd ed. Prentice Hall; 2002.
- [6] Batstone DJ, Keller J, Angelidaki I, Kalyuzhnyi SV, Pavlostathis SG, Rozzi A, et al. The IWA Anaerobic Digestion Model No 1 (ADM1). *Water Sci Technol.* 2002 may;45(10):65–73.
- [7] Bernard O, Hadj-Sadok Z, Dochain D, Genovesi A, Steyer JP. Dynamical model development and parameter identification for an anaerobic wastewater treatment process. *Biotechnol Bioeng.* 2001;75(4):424–438.
- [8] Bastin G, Dochain D. *On-line Estimation and Adaptive Control of Bioreactors.* Elsevier; 1990.
- [9] Yu L, Wensel P, Ma J, Chen S. *Mathematical Modeling in Anaerobic Digestion (AD).* Biorem. 2013;2014:1–12.
- [10] Williams MO, Kevrekidis IG, Rowley CW. A Data-Driven Approximation of the Koopman Operator: Extending Dynamic Mode Decomposition. *Nonlinear Sci.* 2015 dec;25(6):1307–1346.
- [11] Schmid PJ. Dynamic mode decomposition of numerical and experimental data. *Fluid Mech.* 2010 aug;656:5–28.
- [12] Bastin G, Van Impe JF. Nonlinear and Adaptive Control in Biotechnology: A Tutorial. *Eur J Control.* 1995;1(1):37–53.
- [13] Shen S, Premier GC, Guwy A, Dinsdale R. Bifurcation and stability analysis of an anaerobic digestion model. *Nonlinear Dyn.* 2007;48(4):391–408.
- [14] Sbarciog M, Loccufier M, Noldus E. The Estimation of Stability Boundaries for an Anaerobic Digestion System. *IFAC Proc Vol.* 2010;43(6):359–364. 11th IFAC Symposium on Computer Applications in Biotechnology.
- [15] Giesl P, Hafstein S. Review on computational methods for Lyapunov functions. *Discrete Contin Dyn Syst.* 2015;20:2291–2331.
- [16] Mezić I. Spectral Properties of Dynamical Systems, Model Reduction and Decompositions. *Nonlinear Dyn.* 2005;41(1):309–325.

- [17] Lan Y, Mezić I. Linearization in the large of nonlinear systems and Koopman operator spectrum. *Physica D*. 2013;242(1):42–53.
- [18] Mauroy A, Mezić I, Moehlis J. Isostables, isochrons, and Koopman spectrum for the action–angle representation of stable fixed point dynamics. *Physica D*. 2013;261:19–30.
- [19] Mauroy A, Mezić I. Global Stability Analysis Using the Eigenfunctions of the Koopman Operator. *IEEE Trans Autom Control*. 2016 nov;61(11):3356–3369.
- [20] Coayla-Teran EA, Mohammed SEA, Ruffino PRC. Hartman-Grobman theorems along hyperbolic stationary trajectories. *Discrete Contin Dyn Syst*. 2007;17(2):281–292.
- [21] Budišić M, Mohr R, Mezić I. Applied Koopmanism. *Chaos*. 2012;22(4):47510.
- [22] Brunton SL, Brunton BW, Proctor JL, Kutz JN. Koopman invariant subspaces and finite linear representations of nonlinear dynamical systems for control. *PLoS ONE*. 2016;11(2):1–19.
- [23] Garcia-Tenorio C, Tellez-Castro D, Mojica-Nava E, Vande Wouwer A. Analysis of a Class of Hyperbolic Systems via Data-Driven Koopman Operator. In: *International Conference on System Theory, Control and Computing (ICSTCC)*; 2019. p. 566–571.
- [24] Klus S, Koltai P, Schütte C. On the numerical approximation of the Perron-Frobenius and Koopman operator. *Comput Dyn*. 2016;3(1):51–79.
- [25] Korda M, Mezić I. On Convergence of Extended Dynamic Mode Decomposition to the Koopman Operator. *Nonlinear Sci*. 2018 apr;28(2):687–710.
- [26] Sudret B. Global sensitivity analysis using polynomial chaos expansions. *Reliab Eng Syst Saf*. 2008;93(7):964–979.
- [27] Walter E, Pronzato L. Identification of parametric models from experimental data. 1st ed. London: Springer-Verlag London; 1997.
- [28] Neumaier A. Solving Ill-Conditioned and Singular Linear Systems: A Tutorial on Regularization. *SIAM Rev*. 2003;40(3):636–666.
- [29] Rosasco L, De Vito E, Caponnetto A, Piana M, Verri A. Are Loss Functions All the Same? *Neural Comput*. 2004;16(5):1063–1076.
- [30] Konakli K, Sudret B. Polynomial meta-models with canonical low-rank approximations: Numerical insights and comparison to sparse polynomial chaos expansions. *Comput Phys*. 2016 sep;321:1144–1169.
- [31] Konakli K, Sudret B. Reliability analysis of high-dimensional models using low-rank tensor approximations. *Probab Eng Mech*. 2016;46:18–36.
- [32] Sbarciog M, Loccufier M, Vande Wouwer A. On the Optimization of Biogas Production in Anaerobic Digestion Systems*. *IFAC Proc Vol*. 2011;44(1):7150 – 7155. 18th IFAC World Congress.
- [33] Korda M, Mezić I. Linear predictors for nonlinear dynamical systems: Koopman operator meets model predictive control. *Automatica*. 2018;93:149–160.