

# Solving Optimal Control Problems for Monotone Systems Using the Koopman Operator

Aivar Sootla, Guy-Bart Stan and Damien Ernst

**Abstract** Koopman operator theory offers numerous techniques for analysis and control of complex systems. In particular, in this chapter we will argue that for the problem of convergence to an equilibrium, the Koopman operator can be used to take advantage of the geometric properties of controlled systems, thus making the optimal solutions more transparent, and easier to analyse and implement. The motivation for the study of the convergence problem comes from biological applications, where easy-to-implement and easy-to-analyse solutions are of particular value. At the moment, theoretical results have been developed for a class of nonlinear systems called monotone systems. However, the core ideas presented here can be applied heuristically to non-monotone systems. Furthermore, the convergence problem can serve as a building block for solving other control problems such as switching between stable equilibria, or inducing oscillations. These applications are illustrated on biologically inspired numerical examples.

**Key words:** bistable systems, pulse-based control, monotone systems, synthetic and systems biology

---

Aivar Sootla

Department of Engineering Science, University of Oxford, Parks Road, Oxford, OX1 3PJ, United Kingdom e-mail: [aivar.sootla@eng.ox.ac.uk](mailto:aivar.sootla@eng.ox.ac.uk)

Guy-Bart Stan

Department of Bionegineering, Imperial College London, South Kensington Campus, London SW7 2AZ, United Kingdom e-mail: [g.stan@imperial.ac.uk](mailto:g.stan@imperial.ac.uk)

Damien Ernst

Montefiore Institute, University of Liège, Allée de la Découverte 10, B-4000, Liège, Belgium e-mail: [dernst@uliege.be](mailto:dernst@uliege.be)

## 1 Introduction

Research in this chapter was motivated by applications in a rather new field of science called Synthetic Biology, which aims at the design of biological devices for useful purposes, such as cell-based production of valuable chemical compounds (cellular biofactories), detection of toxins or viruses in a sample (whole-cell biosensors), etc. [4, 25, 8]. Engineering novel biological devices is of high interest since the cell machinery opens the door to the synthesis of many useful organic compounds, which cannot be easily realized by other means. This opens avenues to novel applications in biotechnology, bioprocess engineering and cell-based medicine. This introduction contains only the basic information on challenges and advances of control theory applications in the field of Synthetic Biology, and we refer the interested reader to [10] for additional background.

The DNA of living organisms encodes information about their biological functions such as growth, development, replication etc. The DNA is a collection of genes, which are read in the processes called gene expression into molecules such as RNAs, proteins etc., which serve a particular purpose within a cell. Linking the information contained in the DNA to the genetic functions it encodes is not straightforward, since proteins, RNAs, and genes interact with each other, creating complex and often poorly characterized networks. For example, some proteins are known to block or activate the expression of particular genes, while other proteins can bind to each other forming inert complexes. The abundance of complex and poorly characterized systems was one of the reasons why synthetic biology attracted an interest from the feedback control community. A simple example of a feedback control problem in the context of cell biology is forcing the amount of a particular protein in the cell to reach a particular level using exogenous stimuli (i.e., control signals). In control theory, we typically design these signals based on observations of the system (called measurements) to serve a particular purpose (mathematically formulated as minimizing a *pre-defined* objective function). External control of protein levels in microorganisms was shown to be possible in [22, 21, 39], where sensing and actuating are implemented using native or engineered cellular mechanisms. A typical way of performing real-time measurements is fusing a fluorescent protein (for example, mCherry, sfGFP, or Venus) [11] to a protein of interest, and using microscopy to measure the fluorescence levels. Actuating can be performed using chemical, light or temperature induction, where a specific chemical, light wavelength or temperature profile (heat shock), respectively, are used to activate or shut down gene expression and thereby the production of certain proteins [11]. While the research mentioned above was successful in controlling simple biochemical processes at both single cell and population levels, many problems were identified. Quantitative measurement of biological signals, which is of paramount importance for effective control, is still a problem due to current technological limitations. This is because fluorescence measurements depend on the measurement device. Furthermore, optimal control methods, such as dynamic programming, prohibit temporal constraints on control signals. For instance, it is not straightforward to constrain

control signals to be step functions in time, however, such restrictions are often met in biological applications.

Considering the modelling and control challenges delineated above, we looked for other techniques that can be used for control of biological processes. In what follows, we focus on the Koopman operator framework, which is a linear infinite-dimensional representation of nonlinear dynamical systems. Due to its linearity, the Koopman operator allows extending the notions of eigenvalues and eigenvectors (called eigenfunctions in the Koopman framework) to nonlinear systems. Following [19], we consider the Koopman operator on a basin of attraction of an exponentially stable equilibrium. If the slowest Koopman eigenvalue is simple, the corresponding eigenfunction (called the dominant eigenfunction) offers convenient geometric analysis tools [20]. In particular, (under some technical conditions) the trajectories originating on one level set of the dominant eigenfunction reach other level sets simultaneously. Furthermore, the equilibrium lies on the zero level set of the dominant eigenfunction. Therefore, the values of the dominant eigenfunction at a point in the state-space can be used to estimate the convergence time from this point to a neighborhood around the equilibrium [20]. We will elaborate on this property in Section 2.2, but we note that the convergence time (and the values of the dominant eigenfunction) can serve as a qualitative measure of distance to the equilibrium.

As a starting point, we consider the study [34], in which the authors restricted the class of considered systems to be monotone (see [2, 9]) and solved the problem of switching between the equilibria of a bistable system using very simple open-loop control signals. Although monotone systems represent a (theoretically) restrictive class of nonlinear systems, many biological systems can be modeled as monotone systems [27]. While the authors of [34] did not present a mathematical formulation for an optimal control problem, which yields the open-loop control signals computed in [34], the study itself was an indication that such a problem formulation may exist. The Koopman operator approach to the problem was proposed in [33], but an optimal control interpretation was still lacking. In this chapter, we focus on the results from [32, 29], where the problem of convergence to an exponentially stable equilibrium was solved for monotone systems. The main idea in [32, 29] was replacing the Euclidean distance to the equilibrium with the convergence time to a specific level set of the dominant eigenfunction, which can be seen as a qualitative measure of distance to the equilibrium. The new problem formulation also allows casting a dynamic optimization program as a static one, thus significantly simplifying the computations while providing easy-to-implement control signals.

It is worth mentioning that the Koopman operator has been employed in optimal control applications, albeit, in a different context. In [23, 24, 12, 14] a data-based approach was taken, which allowed inferring control laws based on the observed or simulated trajectories. In [17] the convergence problem was formulated using the Koopman operator. While the class of systems was not restricted, full controllability was assumed, i.e. control signals affect all the states in an affine manner. In [40] it was proposed to synchronize cardiac cells by formulating an optimal control problem with the state at the terminal time restricted to a specified level set of

the dominant eigenfunction. However, the authors did not parametrize the control signal, which led to complicated optimal control signals.

The rest of the chapter is organized as follows. In Section 2, we cover basic definitions, pertaining to the Koopman operator and monotone systems theory. We set up the convergence problem in Section 3 on a specific case study and discuss previous research addressing the problem, its limitations and main outcomes. In Section 4, we discuss how to formulate the convergence problem in an easier way so as to provide efficient solutions to the convergence problem. Section 5 is devoted to possible generalizations, while in Section 6 we illustrate the results on numerical examples. We conclude in Section 8.

## 2 Theoretical Background

### 2.1 Dynamical Systems

Consider a system of the form

$$\dot{\mathbf{x}} = \mathbf{F}(\mathbf{x}, u), \quad \mathbf{x}(0) = \mathbf{x}^0, \quad (1)$$

with  $\mathbf{F} : \mathcal{D} \times \mathcal{U} \rightarrow \mathbb{R}^n$ ,  $u : \mathbb{R} \rightarrow \mathcal{U}$ , and where the sets  $\mathcal{D} \subset \mathbb{R}^n$ ,  $\mathcal{U} \subset \mathbb{R}$  are convex, compact, and  $u$  belongs to the space  $\mathcal{U}_\infty$  of Lebesgue measurable functions with values from  $\mathcal{U}$ . We assume that  $\mathbf{F}(\mathbf{x}, u)$  is twice continuously differentiable ( $C^2$ ) in  $(\mathbf{x}, u)$  on an open set containing  $\mathcal{D} \times \mathcal{U}$ . Compactness of  $\mathcal{D}$ ,  $\mathcal{U}$  is required for monotonicity certificates, but convexity can be relaxed [2]. Smoothness of  $F$  is required for Koopman operator analysis [20, 19]. The flow map  $\mathbf{S} : \mathbb{R} \times \mathcal{D} \times \mathcal{U}_\infty \rightarrow \mathbb{R}^n$  induced by the system is such that  $\mathbf{S}(t, \mathbf{x}^0, u)$  is a solution of (1) with an initial condition  $\mathbf{x}^0$  and a control signal  $u$ . We denote the closure of the set  $X$  as  $\text{cl}(X)$ , and its interior as  $\text{int}(X)$ . We denote the differential of a function  $\mathbf{G}(\mathbf{x}, \mathbf{y}) : \mathbb{R}^n \times \mathbb{R}^m \rightarrow \mathbb{R}^k$  with respect to  $\mathbf{x}$  as  $\partial_{\mathbf{x}}\mathbf{G}(\mathbf{x}, \mathbf{y})$ . Let  $\mathbf{J}(\mathbf{x})$  denote the Jacobian matrix of  $\mathbf{F}(\mathbf{x}, 0)$  (i.e.,  $\mathbf{J}(\mathbf{x}) = \partial_{\mathbf{x}}\mathbf{F}(\mathbf{x}, 0)$ ). For an equilibrium  $\mathbf{x}^*$ , the eigenvalues of  $\mathbf{J}(\mathbf{x}^*)$  are  $\{\lambda_1, \dots, \lambda_n\}$  (counted with their algebraic multiplicities) and are ordered by their real part, that is  $\Re(\lambda_i) \geq \Re(\lambda_j)$  for all  $i < j$ . We also assume the eigenvectors of  $\mathbf{J}(\mathbf{x}^*)$  are linearly independent (i.e.,  $\mathbf{J}(\mathbf{x}^*)$  is diagonalizable).

### 2.2 Koopman operator

We denote a semigroup of Koopman operators as  $U^t$ , where  $U^t f(\mathbf{x}) = f \circ \mathbf{S}(t, \mathbf{x})$  for all  $t \geq 0$ , and we restrict the observables to be  $f : \mathbb{R}^n \rightarrow \mathbb{C}$ . We consider only systems with  $\mathbf{F} \in C^2$  around an exponentially stable equilibrium  $\mathbf{x}^*$  (that is, the eigenvalues  $\lambda_j$  of  $\mathbf{J}(\mathbf{x}^*)$  are such that  $\Re(\lambda_j) < 0$  for all  $j$ ) with a diagonalizable  $\mathbf{J}(\mathbf{x}^*)$ . In this case, the eigenvalues  $\lambda_j$  of the Jacobian matrix  $\mathbf{J}(\mathbf{x}^*)$  are also the eigenvalues of the

Koopman operator and are called Koopman eigenvalues, which are associated with the Koopman eigenfunctions  $\phi_{\lambda_j} : \mathcal{C} \rightarrow \mathbb{C}$  satisfying

$$U^t \phi_{\lambda_j}(\mathbf{x}) = \phi_{\lambda_j}(\mathbf{S}(t, \mathbf{x})) = \phi_{\lambda_j}(\mathbf{x}) e^{\lambda_j t}, \quad \mathbf{x} \in \mathcal{C}, \quad (2)$$

where  $\mathcal{C}$  is a subset of  $\mathcal{B}(\mathbf{x}^*) = \{\mathbf{x} \in \mathbb{R}^n \mid \lim_{t \rightarrow \infty} \mathbf{S}(t, \mathbf{x}) = \mathbf{x}^*\}$  – the basin of attraction of  $\mathbf{x}^*$ . Furthermore, Koopman eigenfunctions  $\phi_{\lambda_j}$  belong to  $C^1$  [19]. If  $\mathcal{C}$  is compact, then the infinitesimal generator for the semigroup can be written as  $L = \mathbf{F}^T \nabla$ , and we have the following relation for Koopman eigenfunctions and eigenvalues:

$$\mathbf{F}^T(\mathbf{x}) \nabla \phi_{\lambda_j}(\mathbf{x}) = \lambda_j \phi_{\lambda_j}(\mathbf{x}). \quad (3)$$

We refer to an eigenvalue  $\lambda_1$  satisfying  $\Re(\lambda_1) > \Re(\lambda_j)$  for all  $\lambda_1 \neq \lambda_j$  as *the dominant eigenvalue*. We assume that such an eigenvalue exists (it is the case for monotone systems) and we call the associated eigenfunction  $\phi_{\lambda_1}$  *the dominant eigenfunction*. Under the assumptions presented above, the dominant eigenfunction  $\phi_{\lambda_1}$  can be computed through the Laplace average

$$f_{\lambda_1}^*(\mathbf{x}) = \lim_{T \rightarrow \infty} \frac{1}{T} \int_0^T (f \circ \mathbf{S}(t, \mathbf{x})) e^{-\lambda_1 t} dt, \quad (4)$$

for some  $f \in C^1$  satisfying  $f(\mathbf{x}^*) = 0$  and  $(\nabla f(\mathbf{x}^*))^T \mathbf{v}_1 \neq 0$ , where  $\mathbf{v}_1$  is a right eigenvector of  $\mathbf{J}(\mathbf{x}^*)$  corresponding to  $\lambda_1$ . The Laplace average  $f_{\lambda_1}^*$  is equal to  $\phi_{\lambda_1}(\mathbf{x})$  up to multiplication by a scalar factor. In order to compute  $f_{\lambda_1}^*$ , we can, for example, choose  $f(\mathbf{x}) = \mathbf{w}_1^T(\mathbf{x} - \mathbf{x}^*)$ , where  $\mathbf{w}_1$  is a left eigenvector of  $\mathbf{J}(\mathbf{x}^*)$  corresponding to  $\lambda_1$ . Other eigenfunctions are typically hard to compute using Laplace averages, but they can be estimated using linear algebraic [19] or data-based methods such as dynamic mode decomposition (DMD) (cf. [26, 38]).

The Koopman eigenfunctions capture important geometric properties of the system. In particular, the dominant Koopman eigenfunction  $\phi_{\lambda_1}$  is related to the notion of *isostables* [20]. If the eigenfunction  $\phi_{\lambda_1}$  is  $C^1$ , an isostable  $\partial \mathcal{B}_\alpha$  associated with the value  $\alpha > 0$  is the boundary of the set  $\mathcal{B}_\alpha = \{\mathbf{x} \in \mathbb{R}^n \mid |\phi_{\lambda_1}(\mathbf{x})| \leq \alpha\}$ . It can be shown that the equilibrium  $\mathbf{x}^*$  lies on  $\partial \mathcal{B}_0$  and trajectories with initial conditions on the same isostable  $\partial \mathcal{B}_{\alpha_1}$  reach other isostables  $\partial \mathcal{B}_{\alpha_2}$ , with  $\alpha_2 < \alpha_1$ , after the same time [20]

$$\mathcal{T} = \frac{1}{|\Re(\lambda_1)|} \ln \left( \frac{\alpha_1}{\alpha_2} \right). \quad (5)$$

Hence isostables capture the dominant (or asymptotic) behavior of the unforced system. For instance, in the case of a simple and real  $\lambda_1$ , for example, it can be shown that the trajectories starting from  $\partial \mathcal{B}_\alpha$  share the same asymptotic evolution

$$\mathbf{S}(t, \mathbf{x}) \rightarrow \mathbf{x}^* + \mathbf{v}_1 \alpha e^{\lambda_1 t}, \quad t \rightarrow \infty,$$

where  $\mathbf{x} \in \partial \mathcal{B}_\alpha$ ,  $\mathbf{v}_1$  is a right eigenvector of  $\mathbf{J}(\mathbf{x}^*)$  corresponding to  $\lambda_1$ .

### 2.3 Partial Orders and Monotone Systems

In the monotone systems theory *partial orders*, denoted as  $\succeq$ , are defined using cones  $\mathcal{K}$  in  $\mathbb{R}^n$  under some conditions on  $\mathcal{K}$ . We, however, consider only the partial order induced by a nonnegative orthant  $\mathbb{R}_{\geq 0}^n = \{\mathbf{x} \in \mathbb{R}^n \mid x_i \geq 0, i = 1, \dots, n\}$ , which is defined as follows:  $\mathbf{x} \succeq \mathbf{y}$  if and only if  $\mathbf{x} - \mathbf{y} \in \mathbb{R}_{\geq 0}^n$  (we write  $\mathbf{x} \not\succeq \mathbf{y}$  if the relation  $\mathbf{x} \succeq \mathbf{y}$  does not hold). We will also write  $\mathbf{x} \succ \mathbf{y}$  if  $\mathbf{x} \succeq \mathbf{y}$  and  $\mathbf{x} \neq \mathbf{y}$ , and  $\mathbf{x} \gg \mathbf{y}$  if  $\mathbf{x} - \mathbf{y} \in \mathbb{R}_{> 0}^n = \{\mathbf{x} \in \mathbb{R}^n \mid x_i > 0, i = 1, \dots, n\}$ . A partial order on the space of signals  $u \in \mathcal{U}_\infty$  is defined in a similar manner:  $u \succeq v$  if  $u(t) - v(t) \in \mathbb{R}_{\geq 0}$  for all  $t \geq 0$ . We now introduce concepts that are important for our subsequent discussion. The set  $[\mathbf{x}, \mathbf{y}] = \{\mathbf{z} \in \mathbb{R}^n \mid \mathbf{x} \preceq \mathbf{z} \preceq \mathbf{y}\}$  is called an *order-interval* in the order  $\succeq$ . For a function  $W : \mathbb{R}^n \rightarrow \mathbb{R} \cup \{-\infty, +\infty\}$ , we refer to the set  $\text{dom}(W) = \{\mathbf{x} \in \mathbb{R}^n \mid |W(\mathbf{x})| < \infty\}$  as its *effective domain*. A function  $W : \mathbb{R}^n \rightarrow \mathbb{R} \cup \{-\infty, +\infty\}$  is called *monotone increasing* if  $W(\mathbf{x}) \geq W(\mathbf{y})$  for all  $\mathbf{x} \succeq \mathbf{y}$  on  $\text{dom}(W)$ . We can now define *control monotone systems*.

**Definition 1** *The system  $\dot{\mathbf{x}} = \mathbf{F}(\mathbf{x}, u)$  is called monotone if  $\mathbf{S}(t, \mathbf{x}, u) \preceq \mathbf{S}(t, \mathbf{y}, v)$  for all  $t \geq 0$ , and for all  $\mathbf{x} \preceq \mathbf{y}$ ,  $u \preceq v$ .*

In the case of unforced dynamical systems we will also consider a stricter definition of monotonicity.

**Definition 2** *The unforced system  $\dot{\mathbf{x}} = \mathbf{F}(\mathbf{x}, 0)$  is strongly monotone if it is monotone and  $\mathbf{x} \prec \mathbf{y}$  implies that  $\mathbf{S}(t, \mathbf{x}, 0) \ll \mathbf{S}(t, \mathbf{y}, 0)$  for all  $t > 0$ .*

A certificate for monotonicity is a condition on the system's vector field and can be found in [2]. Monotone systems possess a number of strong stability properties. For example, stable periodic orbits are not possible with  $u = 0$  [9], the small gain theorem can be cast as a condition on the static input-output response of the systems [2]. A rather minor property of monotone systems that we will use is the geometric interpretation of basins of attraction, the proof of which can be found in, e.g. [34].

**Proposition 1** *Let the system  $\dot{\mathbf{x}} = \mathbf{F}(\mathbf{x})$  be monotone on the basin of attraction  $\mathcal{B}(\mathbf{x}^*)$  of an asymptotically stable fixed point  $\mathbf{x}^*$ , then for any  $\mathbf{x}, \mathbf{y}$  in  $\mathcal{B}(\mathbf{x}^*)$  the order interval  $[\mathbf{x}, \mathbf{y}]$  lies in  $\mathcal{B}(\mathbf{x}^*)$ .*

One of the fundamental results for this chapter is the description of spectral properties of monotone systems using the Koopman operator [31].

**Proposition 2** *Assuming that the system  $\dot{\mathbf{x}} = \mathbf{F}(\mathbf{x})$  with  $\mathbf{F} \in C^2$  has an exponentially stable equilibrium  $\mathbf{x}^*$  and that  $\mathbf{J}(\mathbf{x}^*)$  is diagonalizable, let  $\lambda_j$  be the eigenvalues of  $\mathbf{J}(\mathbf{x}^*)$  such that  $\Re(\lambda_i) \geq \Re(\lambda_j)$  for all  $i \leq j$ .*

(i) *If the system is monotone with respect to  $\mathcal{K}$  on a set  $\mathcal{C} \subseteq \mathcal{B}(\mathbf{x}^*)$ , then  $\lambda_1$  is real and negative, the right eigenvector  $\mathbf{v}_1$  of  $\mathbf{J}(\mathbf{x}^*)$  can be chosen such that  $\mathbf{v}_1 \succ 0$ , while the eigenfunction  $\phi_{\lambda_1}$  can be chosen such that  $\phi_{\lambda_1}(\mathbf{x}) \geq \phi_{\lambda_1}(\mathbf{y})$  for all  $\mathbf{x}, \mathbf{y} \in \mathcal{C}$*

satisfying  $\mathbf{x} \succeq \mathbf{y}$ ;

(ii) Furthermore, if the system is strongly monotone with respect to  $\mathcal{K}$  on a set  $\mathcal{C} \subseteq \mathcal{B}(\mathbf{x}^*)$  then  $\lambda_1$  is simple, real and negative,  $\lambda_1 > \Re(\lambda_j)$  for all  $j \geq 2$ ,  $\mathbf{v}_1$  and  $\phi_{\lambda_1}$  can be chosen such that  $\mathbf{v}_1 \gg 0$  and  $\phi_{\lambda_1}(\mathbf{x}) > \phi_{\lambda_1}(\mathbf{y})$  for all  $\mathbf{x}, \mathbf{y} \in \mathcal{C}$  satisfying  $\mathbf{x} \succ \mathbf{y}$ .

Without loss of generality, we will assume that a dominant eigenfunction  $\phi_{\lambda_1}$  is monotone increasing even if  $\lambda_1$  is not simple. In the linear case, this result reduces to the celebrated Perron-Frobenius theorem (cf. [3]). Intuitively, this should not be surprising in light of Koopman operator theory. In [15] it was shown that the Koopman operator provides a global extension of the Hartman-Grobman linearization theorem (cf. [28]) for globally exponentially stable nonlinear systems under sufficient smoothness assumptions.

### 3 Problem Motivation and Issues with Naive Formulations

Due to many challenges in the control of biological systems, it is reasonable to start by solving simple control problems. With this in mind, we formulated the following problem.

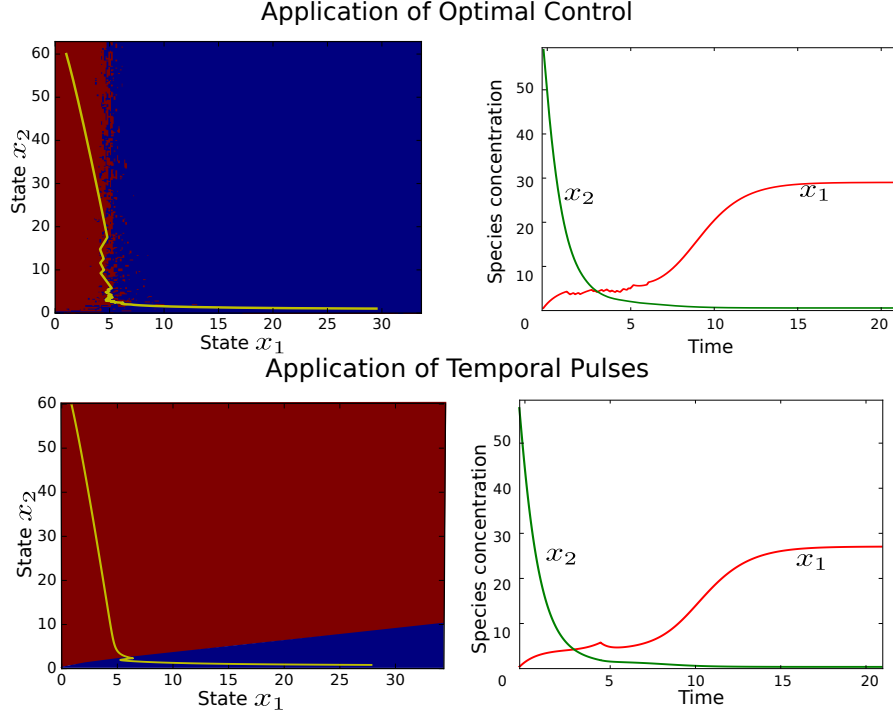
*Problem 1. Converging to an equilibrium.* Consider the system  $\dot{\mathbf{x}} = \mathbf{F}(\mathbf{x}, u)$  and the initial state  $\mathbf{x}^0$ . Compute a control signal  $u(t) \in \mathcal{U}_\infty$  such that the flow  $\mathbf{S}(t, \mathbf{x}^0, u(\cdot))$  reaches an  $\varepsilon$ -ball around  $\mathbf{x}^*$  for some small  $\varepsilon > 0$  in minimum time units  $\mathcal{T}_{\text{conv}}$  subject to the energy budget  $\|u\|_{L_1} \leq \mathcal{E}_{\text{max}}$ .

One straightforward application of the convergence problem is the problem of switching between stable equilibria in multistable systems. As an initial case study, consider a simplified model of a bistable genetic toggle switch system:

$$\begin{aligned} \dot{x}_1 &= p_{11} + \frac{p_{12}}{1 + (x_2)^{p_{13}}} - p_{14}x_1 + bu, \\ \dot{x}_2 &= p_{21} + \frac{p_{22}}{1 + (x_1)^{p_{23}}} - p_{24}x_2, \end{aligned} \tag{6}$$

where  $u$  is assumed to be nonnegative,  $p_{ij}, b$  are nonnegative, and the parameters  $p_{13}, p_{23}$  are typically integers larger than one and smaller than six. For a range of parameter values this model is forward-invariant on  $\mathbb{R}_{\geq 0}^2$  with two exponentially stable equilibria in  $\mathbb{R}_{\geq 0}^2$ , which we denote as  $\mathbf{x}^\bullet$  and  $\mathbf{x}^*$ . In  $\mathbf{x}^*$  the value  $x_1^*$  is much larger than  $x_2^*$ , while in  $\mathbf{x}^\bullet$  the value  $x_2^\bullet$  is much larger than  $x_1^\bullet$ .

In [36],  $\mathbf{x}^*$  was chosen as the target equilibrium and  $\mathbf{x}^\bullet$  as the initial point  $\mathbf{x}^0$ , and the authors opted for an approximate dynamic programming solution. Specifically, they used a batch-mode reinforcement learning approach [5]. In order to apply this method they discretized the model using  $d\mathbf{x}/dt = (\mathbf{x}(k+1) - \mathbf{x}(k))/\Delta$  for some small positive  $\Delta$  and formulated the optimal control problem as:



**Fig. 1** Switching the toggle. We set  $p_{11} = p_{21} = 1$ ,  $p_{12} = 60$ ,  $p_{22} = 30$ ,  $p_{13} = 3$ ,  $p_{23} = 2$ ,  $p_{14} = p_{24} = 1$ ,  $b = 5$ ,  $\Delta = 0.1$ . Top panel: Application of approximate dynamic programming. We formulated the cost with  $\gamma = 0.7$ ,  $\beta = 0.8$ ,  $\|u\| = |u|$ ,  $u \in [0, 1]$ , and  $d(\mathbf{x}, \mathbf{y}) = \log(\max(x_1/y_1, \dots, x_n/y_n) / \min(x_1/y_1, \dots, x_n/y_n))$  as a measure of distance. We ran the algorithm from [5] with  $3 \cdot 10^5$  data points and 70 iterations. In the left panel, the red area indicates the set where  $u = 1$ ; the blue area denotes the set where  $u = 0$ ; and the yellow curve denotes the controlled trajectory. In the right panel, we plot the controlled trajectories versus time. Bottom Panel: Application of temporal pulses. The red and blue areas delineate the basins of attraction of the two equilibria and the yellow curve denotes the trajectory controlled with a temporal pulse with the following characteristics:  $\kappa = 5$ ,  $\tau = 4.7$ . In the right panel, we plot the controlled trajectories versus time.

$$V(\mathbf{x}) = \inf_{u(\cdot)} \sum_{k=0}^{\infty} \gamma^k (d(\mathbf{x}(k), \mathbf{x}^*) + \beta \|u(k)\|),$$

where  $\gamma \in (0, 1)$  is called the “discount factor” and is used to guarantee that the problem is well-posed,  $d(\cdot, \cdot)$  is a distance function in  $\mathbb{R}^n$  and  $\|\cdot\|$  is a norm, while  $\beta$  is a positive scalar. Optimal solutions for different functions  $d(\cdot, \cdot)$ , norms  $\|\cdot\|$  and scalars  $\beta$  were computed. In what follows we will illustrate the results with  $\|\cdot\|$  being the absolute value, while  $d(\cdot, \cdot)$  is chosen to be the Hilbert metric, which provided more robust results to changes in  $\mathbf{x}^*$  than the Euclidean metric. Note that the objective function is only a mathematical representation of the problem we want to solve. In particular, we do not impose a hard constraint on the norm of the control



signals and we require the trajectory to converge to  $\mathbf{x}^*$ . While it is possible to impose the specified constraints theoretically, in practical terms we can pick (by tuning) a large enough  $\beta$ , and ensure that the constraint is satisfied. Furthermore, the penalty on the control signal will ensure that we stop applying control in a vicinity of  $\mathbf{x}^*$ . The main idea of the algorithm in [5] is to use the simulated or observed trajectories of the model in order to solve the optimal control problem. Using the dynamic programming principle the problem is cast as an approximation of the objective function  $V$ . With a growing number of samples in the trajectories, the solution of the approximate dynamic programming converges to a unique solution under some mild conditions [5].

We depict numerical solutions in Figure 1. The optimal control signal forces jittering in the trajectory by constantly switching on and off the control action. This is not an artefact of our approximations: decreasing  $\Delta$  and increasing the number of computed trajectories still yields a jittering behavior in the computed solution. The underlying issue is the singularity of the continuous-time optimal control problem. We will explain this behavior using geometrical arguments. We first note that the control signal will mostly take minimal (0) and maximal ( $u_{\max}$ ) values for this kind of problems following the bang-bang control principle. If the trajectory of the system is in the basin of attraction of  $\mathbf{x}^*$  (denoted as  $\mathcal{B}(\mathbf{x}^*)$ ), then the trajectories will converge to  $\mathbf{x}^*$  even if the control is set to zero. If the trajectories are in  $\mathcal{B}(\mathbf{x}^\bullet)$ , then trajectories will converge to  $\mathbf{x}^\bullet$  in the absence of control signal. If the manifold where the optimal control signal switches between zero and  $u_{\max}$  lies in  $\mathcal{B}(\mathbf{x}^\bullet)$ , then we will observe jittering in the trajectory until it leaves  $\mathcal{B}(\mathbf{x}^\bullet)$ . In order to avoid jittering, one needs to employ singular optimal control methods, however, using this mathematical apparatus seemed rather excessive. Indeed, one can easily show that applying constant control signals is enough to ensure convergence to  $\mathbf{x}^*$  when starting from  $\mathbf{x}^\bullet$ . More generally, we can parametrize control signals as follows:

$$u = \kappa h(t, \tau), \text{ where } h(t, \tau) = \begin{cases} 1 & t \leq \tau \\ 0 & t > \tau \end{cases} \quad (7)$$

Furthermore, if a pair  $(\kappa_1, \tau_1)$  forces the trajectory to move from  $\mathbf{x}^\bullet$  to  $\mathbf{x}^*$  then for any  $\kappa_2 \geq \kappa_1$  and  $\tau_2 \geq \tau_1$ , the pair  $(\kappa_2, \tau_2)$  will also force a switch from  $\mathbf{x}^\bullet$  to  $\mathbf{x}^*$  (6) [34]. This property was used to provide efficient and tight estimates on the set of all  $(\kappa, \tau)$  allowing the switch. Finally, a closed-loop control law (control law depending on the state) can be constructed in a straightforward manner provided we can build an outer approximation of  $\mathcal{B}(\mathbf{x}^\bullet)$  — the basin of attraction of  $\mathbf{x}^\bullet$ . Indeed, in our case  $\mathbb{R}_{\geq 0}^2 \subset \text{cl}(\mathcal{B}(\mathbf{x}^\bullet) \cup \mathcal{B}(\mathbf{x}^*))$ , therefore if  $\mathcal{C}$  contains  $\mathcal{B}(\mathbf{x}^\bullet)$ , then the set  $\mathbb{R}_{\geq 0}^2 / \mathcal{C}$  is contained in  $\mathcal{B}(\mathbf{x}^*)$  and we can define the closed-loop control law as follows:

$$u(\mathbf{x}) = \begin{cases} \kappa & \mathbf{x} \in \mathcal{C}, \\ 0 & \mathbf{x} \in \mathbb{R}_{\geq 0}^2 / \mathcal{C}. \end{cases} \quad (8)$$

This idea is simple and intuitive and can be applied to the class of *monotone* systems [34], to which the system (6) belongs.

The control signal (7) falls into the category of *open-loop control* signals. That is, the control signal does not depend on the states  $\mathbf{x}$ , but only on time, and the control signal is computed in advance and cannot be changed later. Closed-loop solutions, such as solutions arising from approximate dynamic programming, are preferable due to their inherent ability to deal with measurement noise, modeling errors and disturbances. While we also constructed a closed-loop control law (8) and there are efficient methods for estimating basins of attraction for monotone systems [31], this solution does not offer much insight into the “geometry” of the problem. Therefore one of the main outcomes of these studies was not a solution, but a series of questions: considering all the benefits of open-loop controls (7), is it possible to formulate an optimal control problem, which gives (7) (or (8)) as optimal solutions? Can we explicitly use geometric properties of the system to design and tune the control law? Can the control laws be computed efficiently? We will provide answers to these questions for the case of monotone systems with the help of the Koopman operator.

## 4 Convergence to an Isostable Problem

### 4.1 Problem Formulation

The eigenfunctions of the Koopman operator can be used to define a metric, which endows some geometric properties of the system. In our case, we need to use only the dominant eigenfunction, therefore we resort to the use of a (contracting) pseudometric  $d_K(\mathbf{x}, \mathbf{y}) = |\phi_{\lambda_1}(\mathbf{x}) - \phi_{\lambda_1}(\mathbf{y})|$  on a basin of attraction  $\mathcal{B}(\mathbf{x}^*)$ . While this pseudometric does not generally generate compact sublevel sets (especially for monotone systems), we can still use it for approximation of the convergence problem, provided we ensure that  $d_K(\mathbf{x}, \mathbf{y})$  is sufficiently small. Note that if  $d_K(\mathbf{x}, \mathbf{x}^*) \leq \alpha$ , then  $\mathbf{x}$  lies in the set  $\mathcal{B}_\alpha(\mathbf{x}^*)$ , whose boundary is an isostable, and hence we can formulate the problem as follows:

*Problem 1\**. *Convergence to an isostable.* Consider the system  $\dot{\mathbf{x}} = \mathbf{F}(\mathbf{x}, u)$  at the initial state  $\mathbf{x}^0$ . Compute a control signal  $u(t) = \kappa h(t, \tau)$  such that the flow  $\mathbf{S}(t, \mathbf{x}^0, u(\cdot))$  reaches the set  $\mathcal{B}_\varepsilon(\mathbf{x}^*)$  for some small  $\varepsilon > 0$  in minimum time units  $\mathcal{T}_{\text{conv}}$  subject to the energy budget  $\|u\|_{\mathbb{L}_1} \leq \mathcal{E}_{\text{max}}$ .

The main challenge in solving this problem is the parametrization of the control signal. Most of the control methods (such as dynamic programming or Pontryagin’s maximum principle) are not tailored to deal with time parametrized control signals since they rely on the semigroup property of the objective function or the dual variable. Hence it is not entirely clear how to systematically approach this problem using optimal control. Equipped with the Koopman operator framework, we will show how to solve Problem 1\* using a *static optimization program* instead of the dynamic optimization program arising from the optimal control formulation. In or-

der to do so we need to restrict the class of considered systems using the following assumptions.

- A1.** The vector field  $\mathbf{F}(\mathbf{x}, u)$  in (1) is twice continuously differentiable in  $(\mathbf{x}, u)$  on an open set containing  $\mathcal{D} \times \mathcal{U}$ . The unforced system (1) has an exponentially stable equilibrium  $\mathbf{x}^*$  in  $\mathcal{D}$  and  $\mathbf{J}(\mathbf{x}^*)$  is diagonalizable;
- A2.** The system is monotone with respect to  $\mathbb{R}_{\geq 0}^n \times \mathbb{R}$  and for all  $\mathbf{x} \in \mathcal{D}$ ,  $u \in \mathcal{U}_\infty$ , the flow  $\mathbf{S}(t, \mathbf{x}, u(\cdot))$  belongs to  $\mathcal{D}$ ;
- A3.** There exists an eigenfunction  $\phi_{\lambda_1}(\mathbf{x})$  such that  $\nabla \phi_{\lambda_1}(\mathbf{x}) \gg 0$  for all  $\mathbf{x} \in \text{dom}(\phi_{\lambda_1})$ . Furthermore,  $\mathbf{F}(\mathbf{x}, \kappa_1) \succ \mathbf{F}(\mathbf{x}, \kappa_2)$  for all  $\mathbf{x} \in \mathcal{D}$  and  $\kappa_1 > \kappa_2 \geq 0$ .

Assumption A1 guarantees existence and uniqueness of solutions, and the existence of continuously-differentiable ( $C^1$ ) Koopman eigenfunctions around the exponentially stable equilibrium  $\mathbf{x}^*$ . Monotonicity in Assumption A2 is the crucial assumption that requires careful checking. Forward-invariance is relatively straightforward to check: in our case  $\mathcal{U}$  is a compact interval  $[0, u_{\max}]$ , and we only need to make sure that  $\mathcal{D} = \text{cl}(\mathcal{B}(\mathbf{x}^*) \cup \mathcal{B}(\mathbf{x}^\bullet))$  and  $\mathbf{S}(t, \mathbf{x}, u_{\max})$  does not leave  $\mathcal{D}$  for all  $t > 0$ ,  $\mathbf{x} \in \mathcal{D}$ . If Assumptions A1–A2 hold, then we have  $\nabla \phi_{\lambda_1}(\mathbf{x}) \succ 0$  and  $\mathbf{F}(\mathbf{x}, \kappa_1) \succeq \mathbf{F}(\mathbf{x}, \kappa_2)$  for  $\kappa_1 > \kappa_2 \geq 0$ . Hence, Assumption A3 serves as a technical assumption that guarantees uniqueness of solutions to the posed control problem and a certain degree of regularity. We will comment throughout the chapter on the case when Assumption A3 does not hold. Parametrization of control signals in *Problem 1\** can potentially be relaxed, since the only fundamental limitation is to set  $u(t)$  to zero after some time  $\tau$ . We discuss this relaxation in Section 5.2.

## 4.2 Existence of Solutions

Restriction of control signals to temporal pulses will be shown to be beneficial from both a computational and an implementation viewpoint. A possible price to pay for this simplification is the loss of feasibility due to the restriction of the space of control signals. This can potentially occur in multistable systems, however, in the case of bistable monotone systems, a few simple arguments show that the space of temporal pulses is rich enough to solve the convergence problem.

**Proposition 3** *Consider a monotone system  $\dot{\mathbf{x}} = \mathbf{F}(\mathbf{x}, u)$  on  $\mathcal{D} \times \mathcal{U}$  under standard conditions for uniqueness and existence of solutions. If:*

*(i) for  $u = 0$  there are two asymptotically stable equilibria  $\mathbf{x}^*$  and  $\mathbf{x}^\bullet$  in  $\mathcal{D}$  such that  $\mathbf{x}^* \gg \mathbf{x}^\bullet$ ,*

*(ii) for any  $u \in \mathcal{U}_\infty$  the flow  $\mathbf{S}(t, \mathbf{x}, u(\cdot))$  lies in the set  $\mathcal{D} = \text{cl}(\mathcal{B}(\mathbf{x}^*) \cup \mathcal{B}(\mathbf{x}^\bullet))$ ,*

*(iii) for every  $\mathbf{x}$  in a small  $\varepsilon$ -ball (in the Euclidean metric) around  $\mathbf{x}^\bullet$ , there exists  $u^0 \in \mathcal{U}_\infty$  such that  $\mathbf{S}(t, \mathbf{x}, u^0)$  converges to  $\mathbf{x}^*$ .*

*Then there exist  $\kappa$  and  $\tau$  such that  $\mathbf{S}(t, \mathbf{x}, \kappa h(\cdot, \tau))$  converges to  $\mathbf{x}^*$*

*Proof.* If  $\mathbf{x}^0 \in \mathcal{B}(\mathbf{x}^*)$ , then we can choose  $u^0 = 0$ , which is a temporal pulse with  $\tau = 0$ .

Consider now the case when  $\mathbf{x}^0$  lies in the  $\varepsilon$ -ball around  $\mathbf{x}^\bullet$ , and  $\varepsilon$  is small enough such that  $\mathbf{x}^0 \in \mathcal{B}(\mathbf{x}^\bullet)$ . According to the point (iii) there exists a control signal  $u^0 \in \mathcal{U}_\infty$  driving the system from  $\mathbf{x}^0$  to  $\mathbf{x}^*$ . Due to monotonicity, we have  $\mathbf{S}(t, \mathbf{x}^0, u^0) \preceq \mathbf{S}(t, \mathbf{x}^0, \kappa)$ , where  $u^0(t) \leq \kappa$  for (almost) all  $t$ . Since for all  $t > 0$ :

$$\mathbf{S}(t, \mathbf{x}^0, 0) \preceq \mathbf{S}(t, \mathbf{x}^0, u^0) \preceq \mathbf{S}(t, \mathbf{x}^0, \kappa),$$

we have that  $\mathbf{S}(t, \mathbf{x}^0, u^0) \in [\mathbf{S}(t, \mathbf{x}^0, 0), \mathbf{S}(t, \mathbf{x}^0, \kappa)]$ . According to the premise, at some time  $\tau$ , the flow  $\mathbf{S}(\tau, \mathbf{x}^0, u^0)$  will be in the basin of attraction of  $\mathcal{B}(\mathbf{x}^*)$ . Now if  $\mathbf{S}(\tau, \mathbf{x}^0, \kappa) \in \mathcal{B}(\mathbf{x}^*)$  we have a contradiction, since  $\mathbf{S}(\tau, \mathbf{x}^0, u^0)$  belongs to  $[\mathbf{S}(\tau, \mathbf{x}^0, 0), \mathbf{S}(\tau, \mathbf{x}^0, \kappa)]$ , which is a subset of  $\mathcal{B}(\mathbf{x}^*)$  according to Proposition 1. Hence if we can switch from  $\mathbf{x}^0$  to  $\mathbf{x}^*$  with a control signal  $u(t)$ , then we can switch with a temporal pulse  $\kappa h(t, \tau)$ .

The case when  $\mathbf{x}^0$  lies in  $\mathcal{B}(\mathbf{x}^*)$ , but not in an  $\varepsilon$  ball around  $\mathbf{x}^\bullet$ , is treated in a similar manner by first allowing the trajectory to converge to a neighborhood of  $\mathbf{x}^0$  with  $u^1 = 0$  and then applying the argument above.  $\square$

The condition  $\mathbf{x}^* \gg \mathbf{x}^\bullet$  (or  $\mathbf{x}^* \ll \mathbf{x}^\bullet$  which can be cast in the form above by changing the order to one induced by  $-\mathbb{R}_{\geq 0}^n$ ) is typically fulfilled for a large class of bistable monotone systems. Condition (iii) is a controllability condition, which ensures the existence of solutions for non-parametrized control signals. Condition (ii) is quite easy to verify as discussed above. Note that in Proposition 3 we relax Assumptions A1 and A2 and this result is valid for a more general class of systems.

We can strengthen the argument for using constant control signals by showing that *constant controls are optimal* in the absence of energy constraints. Consider the following optimal control problem over bounded and measurable control signals:

$$\begin{aligned} V(\mathbf{z}, \kappa, \beta) &= \inf_{\tau, u \in \mathcal{U}_\infty([0, \kappa])} \tau, & (9) \\ \text{subject to (1), } & \mathbf{x}(0) = \mathbf{z}, \\ & \mathbf{x}(\tau) \in \mathcal{C}_\beta = \{\mathbf{y} \in \mathbb{R}^n \mid \phi_{\lambda_1}(\mathbf{y}) = \beta\}, \end{aligned}$$

where  $\phi_{\lambda_1}(\mathbf{x})$  is a  $C^1$  increasing dominant eigenfunction defined on the basin of attraction of  $\mathbf{x}^*$ . Under our assumptions, the solution to this problem is surprisingly straightforward.

**Proposition 4** *Let the system (1) satisfy Assumptions A1 – A3. Then*

(i) *If  $\phi_{\lambda_1}(\mathbf{z}) < \beta$ , then the optimal solution to (9), if it exists, is  $u(t) = \kappa$  for all  $t \in [0, \tau]$ ;*

(ii) *If  $\phi_{\lambda_1}(\mathbf{z}) \geq \beta$ , then the optimal solution to (9) is  $u(t) = 0$  for all  $t \geq 0$ .*

*Proof.* (i) Let  $u^0(t) = \kappa$  for all  $t > 0$ , and  $u^\delta(t)$  be any admissible control signal. Then  $u^0(t) \succeq u^\delta(t)$  for all  $t \in [0, \tau]$ . By monotonicity we then have  $\mathbf{S}(t, \mathbf{z}, u^0(\cdot)) \succeq \mathbf{S}(t, \mathbf{z}, u^\delta(\cdot))$ , which leads to  $\phi_{\lambda_1}(\mathbf{S}(t, \mathbf{z}, u^0(\cdot))) \geq \phi_{\lambda_1}(\mathbf{S}(t, \mathbf{z}, u^\delta(\cdot)))$  for all  $t \geq 0$  due to Proposition 2. Hence

$$\begin{aligned}\beta &= \phi_{\lambda_1}(\mathbf{S}(\tau, \mathbf{z}, u^0(\cdot))) \geq \phi_{\lambda_1}(\mathbf{S}(\tau, \mathbf{z}, u^\delta(\cdot))), \\ \beta &> \phi_{\lambda_1}(\mathbf{S}(t, \mathbf{z}, u^0(\cdot))) \geq \phi_{\lambda_1}(\mathbf{S}(t, \mathbf{z}, u^\delta(\cdot))) \text{ for } t < \tau,\end{aligned}$$

which implies that the target set  $\mathcal{C}_\beta$  is reached with  $u^0(\cdot)$  at least as fast as with any other admissible control signal  $u^\delta(\cdot)$ . Therefore, the control signal  $u^0(t) = \kappa$  is an optimal solution of the problem.

(ii) The proof is similar to the point (i).  $\square$

### 4.3 Pulse Control Function

The computational solution to *Problem 1\** relies on a function of state  $\mathbf{x}$ , of the pulse magnitude  $\kappa$  and of the pulse length  $\tau$ , which we call *the pulse control function*.

**Definition 3** Let the function  $r : \mathcal{D} \times \mathbb{R}_{\geq 0} \times \mathbb{R}_{\geq 0} \rightarrow \mathbb{C} \cup \{\infty\}$  such that

$$r(\mathbf{x}, \kappa, \tau) = \phi_{\lambda_1}(\mathbf{S}(\tau, \mathbf{x}, \kappa)),$$

where  $\phi_{\lambda_1}$  is a dominant eigenfunction on the basin of attraction of  $\mathbf{x}^*$ , be called the pulse control function. By convention  $r(\mathbf{x}, \kappa, \tau) = \infty$ , if  $\mathbf{S}(\tau, \mathbf{x}, \kappa) \notin \mathcal{B}(\mathbf{x}^*)$ .

If  $\phi_{\lambda_1}$  is real-valued and monotone increasing on  $\text{dom}(\phi_{\lambda_1}) = \mathcal{B}(\mathbf{x}^*)$ , then we assume it is extended to  $\mathcal{D}$  so that  $\phi_{\lambda_1} : \mathcal{D} \rightarrow \mathbb{R} \cup \{-\infty, +\infty\}$  is monotone increasing on  $\mathcal{D}$ . We note that the case when the point  $\mathbf{x}$  is another equilibrium  $\mathbf{x}^\bullet$  was considered in [33]. Estimation of the pulse control function also relies on a Laplace average (4), similarly to what was done for the estimation of  $\phi_{\lambda_1}$ :

$$\begin{aligned}r(\mathbf{x}, \kappa, \tau) &= \lim_{\bar{t} \rightarrow \infty} \frac{1}{\bar{t}} \int_0^{\bar{t}} f \circ \mathbf{S}(t, \mathbf{S}(\tau, \mathbf{x}, \kappa), 0) e^{-\lambda_1 t} dt \\ &= \lim_{\bar{t} \rightarrow \infty} \frac{1}{\bar{t}} \int_\tau^{\bar{t}} f \circ \mathbf{S}(t, \mathbf{x}, \kappa h(\cdot, \tau)) e^{-\lambda_1(t-\tau)} dt, \quad (10)\end{aligned}$$

where  $\lambda_1$  is the dominant eigenvalue of  $\mathbf{J}(\mathbf{x}^*)$  with a corresponding right eigenvector  $\mathbf{v}_1$ ,  $f \in C^1$  satisfies  $f(\mathbf{x}^*) = 0$ ,  $\mathbf{v}_1^T \nabla f(\mathbf{x}^*) \neq 0$ , and  $h(t, \tau)$  is the step function defined in (7). In practice, we choose  $f(\mathbf{x}) = \mathbf{w}_1^T (\mathbf{x} - \mathbf{x}^*)$ , where  $\mathbf{w}_1$  is the dominant left eigenvector of  $\mathbf{J}(\mathbf{x}^*)$ . If Assumptions A1 and A2 hold then  $\lambda_1$  is real according to Proposition 2), and we have

$$r(\mathbf{x}, \kappa, \tau) \approx \mathbf{w}_1^T (\mathbf{S}(\bar{t}, \mathbf{x}, \kappa h(\cdot, \tau)) - \mathbf{x}^*) e^{-\lambda_1(\bar{t}-\tau)}$$

where the time  $\bar{t}$  should be chosen large enough. In particular, the value of  $\bar{t}$  should be chosen such that  $\mathbf{w}_1^T (\mathbf{S}(t, \mathbf{x}, \kappa h(\cdot, \tau)) - \mathbf{x}^*) e^{-\lambda_1(t-\tau)}$  converges to a constant value. Naturally, with  $\bar{t}$  tending to infinity the numerical integration errors can lead to di-

vergence of  $\mathbf{w}_1^T(\mathbf{S}(t, \mathbf{x}, \kappa h(\cdot, \tau)) - \mathbf{x}^*)e^{-\lambda_1(t-\tau)}$ . Therefore, the tolerance of the differential equation solver should be set to  $O(e^{\lambda_1(\bar{t}-\tau)})$ .

For monotone systems the pulse control function  $r$  possesses strong properties, which stem from Proposition 2 and are the key element for this approach.

**Lemma 1** *Let the system (1) satisfy Assumptions A1–A3. Then  $r$  is a  $C^1$  function on its effective domain  $\text{dom}(r)$ . Furthermore, for all  $(\mathbf{x}, \kappa, \tau) \in \text{dom}(r)$*

- (i)  $\partial_{\mathbf{x}}r(\mathbf{x}, \kappa, \tau) \gg 0$ ,  $\partial_{\kappa}r(\mathbf{x}, \kappa, \tau) > 0$  and  $\partial_{\tau}r(\mathbf{x}, \kappa, \tau) > \lambda_1 r(\mathbf{x}, \kappa, \tau)$ ;
- (ii) *If  $r(\mathbf{x}, \kappa, \tau) \leq 0$ , then  $\partial_{\tau}r(\mathbf{x}, \kappa, \tau) > 0$ .*

*Proof.* (o) First, we need to show that under the assumptions above for all  $t > 0$  and  $\kappa_1 > \kappa_2$ , we have

$$\mathbf{S}(t, \mathbf{x}, \kappa_1) \succ \mathbf{S}(t, \mathbf{x}, \kappa_2), \quad (11)$$

$$\phi_{\lambda_1}(\mathbf{S}(t, \mathbf{x}, \kappa_1)) > \phi_{\lambda_1}(\mathbf{S}(t, \mathbf{x}, \kappa_2)). \quad (12)$$

Since  $\mathbf{S}(t, \mathbf{x}, \kappa_1) \succeq \mathbf{S}(t, \mathbf{x}, \kappa_2)$  for all  $t > 0$  and  $\kappa_1 > \kappa_2$  due to Assumption A2 (monotonicity), we only need to show that  $\mathbf{S}(t, \mathbf{x}, \kappa_1) \neq \mathbf{S}(t, \mathbf{x}, \kappa_2)$  for all finite  $t > 0$ . At  $t = 0$ , the time derivatives of the flows  $\mathbf{S}(t, \mathbf{x}, \kappa_1)$  and  $\mathbf{S}(t, \mathbf{x}, \kappa_2)$  are equal to  $\mathbf{F}(\mathbf{x}, \kappa_1)$  and  $\mathbf{F}(\mathbf{x}, \kappa_2)$ , respectively. Due to Assumption A3 we have that  $\mathbf{F}(\mathbf{x}, \kappa_1) \succ \mathbf{F}(\mathbf{x}, \kappa_2)$  and consequently there exists a  $T > 0$  such that  $\mathbf{S}(t, \mathbf{x}, \kappa_1) \succ \mathbf{S}(t, \mathbf{x}, \kappa_2)$  for all  $t < T$ . If for some  $\xi \geq T$  we have that  $\mathbf{S}(\xi, \mathbf{x}, \kappa_1) = \mathbf{S}(\xi, \mathbf{x}, \kappa_2)$  and  $\mathbf{S}(\xi, \mathbf{x}, \kappa_1) \succ \mathbf{S}(\xi, \mathbf{x}, \kappa_2)$  for all  $t < \xi$ , then for some index  $i$  we have

$$\left. \frac{d\mathbf{S}_i(t, \mathbf{x}, \kappa_1)}{dt} \right|_{t=\xi} < \left. \frac{d\mathbf{S}_i(t, \mathbf{x}, \kappa_2)}{dt} \right|_{t=\xi}.$$

This leads to  $F_i(\mathbf{S}(\xi, \mathbf{x}, \kappa_1), \kappa_1) < F_i(\mathbf{S}(\xi, \mathbf{x}, \kappa_2), \kappa_2)$ , which together with  $\mathbf{S}(\xi, \mathbf{x}, \kappa_1) = \mathbf{S}(\xi, \mathbf{x}, \kappa_2)$  contradicts Assumption A3. Therefore,  $\mathbf{S}(t, \mathbf{x}, \kappa_1) \succ \mathbf{S}(t, \mathbf{x}, \kappa_2)$  for all finite  $t > 0$ . Due to Assumption A3 we also have that  $\nabla\phi_{\lambda_1}(\mathbf{x}) \gg 0$ , which in particular means that  $\phi_{\lambda_1}(\mathbf{x}) > \phi_{\lambda_1}(\mathbf{y})$  for all  $\mathbf{x} \succ \mathbf{y}$ , and (12) follows from (11).

(i) Assumption A1 guarantees that  $\mathbf{F}(\mathbf{x}, u) \in C^2$  and hence the flow  $\mathbf{S}(t, \mathbf{x}, u)$  is continuously-differentiable for constant control signals. Combining this with the fact that  $\phi_{\lambda_1} \in C^1$  it follows that  $r(\mathbf{x}, \kappa, \tau) = \phi_{\lambda_1}(\mathbf{S}(\tau, \mathbf{x}, \kappa))$  is a  $C^1$  function.

For  $(\mathbf{x}, \kappa, \tau)$  and  $(\mathbf{y}, \kappa, \tau) \in \text{dom}(r)$  such that  $\mathbf{x} \succ \mathbf{y}$ , monotonicity and  $\nabla\phi_{\lambda_1}(\mathbf{x}) \gg 0$  (Assumption A3) ensure that  $\phi_{\lambda_1}(\mathbf{S}(\tau, \mathbf{x}, \kappa)) > \phi_{\lambda_1}(\mathbf{S}(\tau, \mathbf{y}, \kappa))$ , which implies that  $\nabla_{\mathbf{x}}r(\mathbf{x}, \kappa, \tau) \gg 0$ .

For  $(\mathbf{x}, \kappa, \tau)$  and  $(\mathbf{x}, \nu, \tau) \in \text{dom}(r)$  such that  $\kappa > \nu$ , monotonicity and point (o) ensure that (12) holds, which implies that  $\partial_{\kappa}r(\mathbf{x}, \kappa, \tau) > 0$ .

Finally, the following derivation leads to  $\partial_{\tau}r(\mathbf{x}, \kappa, \tau) > \lambda_1 r(\mathbf{x}, \kappa, \tau)$ :

$$\begin{aligned} \partial_\tau r(\mathbf{x}, \boldsymbol{\kappa}, \tau) &= \left. \frac{d\phi_{\lambda_1}(\mathbf{S}(t, \mathbf{x}, \boldsymbol{\kappa}))}{dt} \right|_{t=\tau} = \\ & \quad \nabla\phi_{\lambda_1}(\mathbf{S}(\tau, \mathbf{x}, \boldsymbol{\kappa}))^T \mathbf{F}(\mathbf{S}(\tau, \mathbf{x}, \boldsymbol{\kappa}), \boldsymbol{\kappa}) >^* \\ & \quad \nabla\phi_{\lambda_1}(\mathbf{S}(\tau, \mathbf{x}, \boldsymbol{\kappa}))^T \mathbf{F}(\mathbf{S}(\tau, \mathbf{x}, \boldsymbol{\kappa}), 0) =^\dagger \\ & \quad \lambda_1 \phi_{\lambda_1}(\mathbf{S}(\tau, \mathbf{x}, \boldsymbol{\kappa})) = \lambda_1 r(\mathbf{x}, \boldsymbol{\kappa}, \tau), \end{aligned}$$

where the inequality  $*$  is due to Assumption A3, and the equality  $\dagger$  is due to (3).

(ii) follows directly from point (i).  $\square$

If Assumption A3 does not hold, then all the inequalities in Lemma 1 are not strict. For instance, we have that  $\partial_{\mathbf{x}} r(\mathbf{x}, \boldsymbol{\kappa}, \tau) \succeq 0$ ,  $\partial_{\boldsymbol{\kappa}} r(\mathbf{x}, \boldsymbol{\kappa}, \tau) \geq 0$  and  $\partial_\tau r(\mathbf{x}, \boldsymbol{\kappa}, \tau) \geq \lambda_1 r(\mathbf{x}, \boldsymbol{\kappa}, \tau)$  in point (i).

In light of Lemma (1), the pulse control function  $r$  has a direct relation to the optimization problem formulated in (9). In particular,  $\phi_{\lambda_1}(\mathbf{x}) < \beta$  implies that  $r(\mathbf{x}, \boldsymbol{\kappa}, V(\mathbf{x}, \boldsymbol{\kappa}, \beta)) = \beta$  provided that the optimization problem in (9) has a solution. On the other hand, there might be some values  $\tau \neq V(\mathbf{x}, \boldsymbol{\kappa}, \beta)$  such that  $r(\mathbf{x}, \boldsymbol{\kappa}, \tau) = \beta$ . In general, if  $\phi_{\lambda_1}(\mathbf{x}) < \beta$ , then

$$V(\mathbf{x}, \boldsymbol{\kappa}, \beta) = \min\{\tau \in \mathbb{R}_{>0} \mid r(\mathbf{x}, \boldsymbol{\kappa}, \tau) = \beta\}, \quad (13)$$

provided that the solution exists. In particular, if  $r$  is strictly monotone increasing in  $\tau$  on some subset of  $(\mathbf{x}, \boldsymbol{\kappa}, \tau)$ , then  $V(\mathbf{x}, \boldsymbol{\kappa}, \beta) = \tau$  if and only if  $r(\mathbf{x}, \boldsymbol{\kappa}, \tau) = \beta$ , which again justifies our use of the pulse control function in these cases.

If the premise of the point (ii) in Lemma 1 holds, then the level sets of  $r$  are the graphs of strictly decreasing functions, a result which simplifies their computations as discussed in [33]. The algorithms developed in [31], [30] and [13] can be used to estimate these level sets.

**Corollary 1** *Let the system (1) satisfy Assumptions A1–A3, and  $\alpha > 0$ . If  $\phi_{\lambda_1}(\mathbf{x}) < -\alpha$ , then  $\{(\boldsymbol{\kappa}, \tau) \in \mathbb{R}_{>0}^2 \mid r(\mathbf{x}, \boldsymbol{\kappa}, \tau) = -\alpha\}$  is the graph of a strictly decreasing function, i.e., this set does not contain pairs  $(\boldsymbol{\kappa}_1, \tau_1) \neq (\boldsymbol{\kappa}_2, \tau_2)$  such that  $\boldsymbol{\kappa}_1 \leq \boldsymbol{\kappa}_2$  and  $\tau_1 \leq \tau_2$ .*

*Proof.* Due to Lemma 1 we have that  $\partial_\tau r(\mathbf{x}, \boldsymbol{\kappa}, \tau) > 0$  as long as  $r(\mathbf{x}, \boldsymbol{\kappa}, \tau) \leq 0$ , and that  $\partial_{\boldsymbol{\kappa}} r(\mathbf{x}, \boldsymbol{\kappa}, \tau) > 0$ . Now since  $\partial_\tau r(\mathbf{x}, \boldsymbol{\kappa}, \tau) > 0$  (for all  $(\boldsymbol{\kappa}, \tau)$  such that  $r(\mathbf{x}, \boldsymbol{\kappa}, \tau) = -\alpha < 0$ ), the implicit function theorem implies that there exists a function  $\tau = g(\boldsymbol{\kappa}, \alpha)$  such that  $r(\mathbf{x}, \boldsymbol{\kappa}, g(\boldsymbol{\kappa}, \alpha)) = -\alpha$ . Furthermore, the function  $g$  is  $C^1$  in  $\boldsymbol{\kappa}$  and  $\partial_{\boldsymbol{\kappa}} g = -\partial_{\boldsymbol{\kappa}} r(\mathbf{x}, \boldsymbol{\kappa}, \tau) / \partial_\tau r(\mathbf{x}, \boldsymbol{\kappa}, \tau) < 0$  in the neighborhood of the level set  $\{(\boldsymbol{\kappa}, \tau) \mid r(\mathbf{x}, \boldsymbol{\kappa}, \tau) = -\alpha\}$ . Therefore the level set  $\{\tau, \boldsymbol{\kappa} \mid r(\mathbf{x}, \boldsymbol{\kappa}, \tau) = -\alpha\}$  is the graph of a strictly decreasing function in  $\boldsymbol{\kappa}$ . It also directly follows that the level set  $\{(\boldsymbol{\kappa}, \tau) \mid r(\mathbf{x}, \boldsymbol{\kappa}, \tau) = -\alpha\}$  is the graph of a strictly decreasing function in  $\tau$ .  $\square$

#### 4.4 Solution to the Problem of Convergence to Isostables

We can now reformulate *Problem 1\** using the pulse control function. If  $\phi_{\lambda_1}(\mathbf{x}^0) \geq \varepsilon$ , then according to Proposition 4 the optimal convergence to the level set  $\phi_{\lambda_1}(\mathbf{x}) = -\varepsilon$  is solved by a zero control signal. Therefore, we can assume that  $\phi_{\lambda_1}(\mathbf{x}^0) \leq -\varepsilon$ , that is we approach the equilibrium  $\mathbf{x}^*$  from below in the partial order. Since  $r(\mathbf{x}, \kappa, \tau) = -\varepsilon$ ,  $\mathbf{x}$  belongs to  $\{\mathbf{x} \in \mathbb{R}^n \mid \phi_{\lambda_1}(\mathbf{S}(\tau, \mathbf{x}, \kappa)) = -\varepsilon\}$  and as a consequence  $\mathbf{x} \in \mathcal{B}_\varepsilon$ . Therefore, the terminal constraint becomes simply  $r(\mathbf{x}, \kappa, \tau) \leq -\varepsilon$ . The convergence time formula is computed using Koopman operator theory as:

$$\mathcal{T}_{\text{conv}} = \frac{1}{|\Re(\lambda_1)|} \ln \left( \frac{|\phi_{\lambda_1}(\mathbf{S}(\tau, \mathbf{x}, \kappa))|}{\varepsilon} \right) + \tau = \frac{1}{|\Re(\lambda_1)|} \ln \left( \frac{|r(\mathbf{x}, \kappa, \tau)|}{\varepsilon} \right) + \tau,$$

where the total convergence time is computed as the sum of the time of free motion of the flow (the first term in the sum) and the time  $\tau$  during which the control is applied (the second term in the sum). Finally, the energy budget constraint in  $\mathbb{L}_1$  norm is  $\kappa \cdot \tau \leq \mathcal{E}_{\text{max}}$  and *Problem 1\** becomes a *static* optimization program:

$$\gamma^* = \min_{\kappa \geq 0, \tau \geq 0} \frac{1}{|\Re(\lambda_1)|} \ln(|r(\mathbf{x}, \kappa, \tau)|) + \tau, \quad (14)$$

$$\text{subject to: } r(\mathbf{x}, \kappa, \tau) \leq -\varepsilon, \quad (15)$$

$$\kappa \cdot \tau \leq \mathcal{E}_{\text{max}}, \quad (16)$$

where in the objective function we removed the constant terms from  $\mathcal{T}_{\text{conv}}$ , that is  $\gamma^* = \mathcal{T}_{\text{conv}} + \frac{1}{|\Re(\lambda_1)|} \ln(\varepsilon)$ .

**Theorem 1** Consider the system (1), *Problem 1\** under Assumptions A1–A3 and the optimization program (14). If  $\phi_{\lambda_1}(\mathbf{x}^0) \leq -\varepsilon$ , an optimal solution to (14) is an optimal solution to *Problem 1\**, if the former is feasible. Furthermore, the objective is nonincreasing in  $\kappa$  and  $\tau$  and an optimal solution to (14), if it exists, is achieved at the boundary of the admissible set to the constraint (15) and/or the constraint (16).

*Proof.* It is straightforward to verify that all the constraints and optimization objectives are the same for *Problem 1* and problem (14). Hence, by construction, the first part of the statement is fulfilled.

Since the system is monotone,  $\lambda_1$  is real and therefore  $\lambda_1 = \Re(\lambda_1)$ . Now according to the constraint (15), we have that  $r(\mathbf{x}^0, \kappa, \tau) < 0$ , which implies the following derivation:

$$\begin{aligned} \partial_\tau (\ln(|r(\mathbf{x}^0, \kappa, \tau)| e^{|\lambda_1| \tau})) &= \frac{\partial_\tau (|r(\mathbf{x}^0, \kappa, \tau)| e^{|\lambda_1| \tau})}{|r(\mathbf{x}^0, \kappa, \tau)| e^{|\lambda_1| \tau}} = \\ &= \frac{-\partial_\tau (r(\mathbf{x}^0, \kappa, \tau)) \cdot e^{|\lambda_1| \tau} + |\lambda_1| |r(\mathbf{x}^0, \kappa, \tau)| e^{|\lambda_1| \tau}}{|r(\mathbf{x}^0, \kappa, \tau)| e^{|\lambda_1| \tau}} <^* \\ &= \frac{\lambda_1 |r(\mathbf{x}^0, \kappa, \tau)| e^{(\lambda_1) \tau} + |\lambda_1| |r(\mathbf{x}^0, \kappa, \tau)| e^{|\lambda_1| \tau}}{|r(\mathbf{x}^0, \kappa, \tau)| e^{|\lambda_1| \tau}} = 0 \end{aligned}$$



where the inequality \* follows from Lemma 1. Hence, the derivative of the objective function in (14) with respect to  $\tau$  is negative. Also,  $\partial_{\kappa} \ln(|r(\mathbf{x}^0, \kappa, \tau)|e^{|\lambda_1|\tau})$  is also negative according to Lemma 1. This shows that if there is a feasible point, the constraints (15), (16) are reached in order to minimize the objective.  $\square$

If Assumption A3 does not hold, then the function  $r$  is not strictly monotone increasing and multiple minima in (14) are possible, including the points which do not activate the constraints, but we can still compute a minimizing solution using the same program. If the premise of Theorem 1 holds then the optimization program can be solved by a line search over  $\kappa$  (or  $\tau$ ) over the constraints boundaries, since the minimum is attained when one of the constraints is active. The two terms in the sum of the objective function (14) establish the tradeoff on the choice of the intermediate target isostable (which is to be reached after a time  $\tau$ ) and the convergence time to the target isostable  $\mathcal{B}_{\varepsilon}(\mathbf{x}^*)$ . For instance, choosing an isostable close to the equilibrium can lead to a large pulse duration  $\tau$  (second term), but a small convergence time of the free motion (first term). Visualisation of the level sets of the function  $r$  and the boundaries of the constraints also allows to understand the tradeoff between the constraints (energy spent) and the objective (convergence time), which is not straightforward using standard optimal control theory. To summarize, we derived a *static optimization problem*, which has the same solution as the *dynamic optimization problem* (Problem 1\* under Assumptions A1–A3).

In the case of the switching problem, that is, converging from another equilibrium  $\mathbf{x}^{\bullet}$  to the target equilibrium  $\mathbf{x}^*$ , we can make a connection to the open loop solution from [34] and [33]. In particular, in [34] the authors compute the effective domain of the function  $r(\mathbf{x}^{\bullet}, \kappa, \tau)$  or the set  $\{(\kappa, \tau) \in \mathbb{R}^2 \mid |r(\mathbf{x}^{\bullet}, \kappa, \tau)| < \infty\}$ . The open loop solutions  $(\kappa, \tau)$  are then picked from this set. In [33], the authors compute the level sets of the function  $r(\mathbf{x}^{\bullet}, \kappa, \tau)$  and the convergence time from  $\mathbf{x}^{\bullet}$ .

## 5 Generalizations

### 5.1 Dealing with Parametric Uncertainty

One of the advantages of monotone systems appears when dealing with parametric uncertainty, provided the system is monotone with respect to parameter changes as well. Consider the system

$$\dot{\mathbf{x}} = \mathbf{F}(\mathbf{x}, \mathbf{p}, u), \quad \mathbf{x}(0) = \mathbf{x}^0, \quad (17)$$

where  $\mathbf{p}$  belongs to a compact set  $\mathcal{P} \in \mathbb{R}^l$ , the flow of the system is  $\mathbf{S} : \mathbb{R}_{\geq 0} \times \mathcal{D} \times \mathcal{P} \times \mathcal{U}_{\infty} \rightarrow \mathcal{D}$ , and for every  $\mathbf{p}$  the system (17) satisfies Assumptions A1–A3. Monotonicity with respect to the parameter  $\mathbf{p}$  can be treated as monotonicity with respect to constant input signals. Monotonicity with respect to parameters allows to study a family of systems using upper and lower bounding systems. For instance, if

$\mathbf{p} \in [\mathbf{p}_1, \mathbf{p}_2] \subseteq \mathcal{P}$ , then for all  $\mathbf{x}$ ,  $u$ , and  $t \geq 0$  we have

$$\mathbf{S}(t, \mathbf{x}, \mathbf{p}_1, u(\cdot)) \preceq \mathbf{S}(t, \mathbf{x}, \mathbf{p}, u(\cdot)) \preceq \mathbf{S}(t, \mathbf{x}, \mathbf{p}_2, u(\cdot)).$$

In biological applications, this property is particularly useful since precise values of parameters are hard to estimate, but intervals of realistic parameter values are readily accessible.

In this section, we discuss the results from [29]. First an auxiliary problem is formulated:

$$V(\mathbf{z}, \kappa, \beta, \mathbf{p}) = \inf_{\tau, u \in \mathcal{U}_\infty([0, \kappa])} \tau, \quad (18)$$

$$\text{subject to: } \dot{\mathbf{x}} = \mathbf{F}(\mathbf{x}, \mathbf{p}, u), \mathbf{x}(0) = \mathbf{z}, \mathbf{x}(\tau) \in \mathcal{A}_{\text{target}} = \{\tilde{\mathbf{x}} \in \mathbb{R}^n \mid g(\tilde{\mathbf{x}}) = \beta\},$$

where  $g \in C^1$  and  $\nabla g \gg 0$  on  $\text{dom}(g)$ . In comparison to (9) there is one major difference: the target set is parametrized by an arbitrary monotone increasing function  $g$ . Since the objective function depends on parameter  $\mathbf{p}$ , parameter-dependence is not taken into account in this formulation. The solution to the problem (18) can be obtained similarly to Proposition 4, that is if  $g(\mathbf{z}) < \beta$  (respectively,  $g(\mathbf{z}) \geq \beta$ ), then  $u^0(\cdot) = \kappa$  (respectively,  $u^0(\cdot) = 0$ ) is an optimal solution of (18). Using this observation, it is possible to derive a bound on  $V$  given parametric uncertainties. In particular, if the system (17) is monotone with respect to parameter variations then

$$V(\mathbf{z}, \kappa, \beta, \mathbf{p}_1) \geq V(\mathbf{z}, \kappa, \beta, \mathbf{p}) \geq V(\mathbf{z}, \kappa, \beta, \mathbf{p}_2),$$

for all  $\mathbf{p} \in [\mathbf{p}_1, \mathbf{p}_2]$  under some technical conditions. The objective functions  $V$  can be computed using an extended pulse control function. To that effect, we simply introduce  $\tilde{r}(\mathbf{x}, \kappa, \tau, \mathbf{p}) = \phi_{\lambda_1, \mathbf{p}}(\mathbf{S}(\tau, \mathbf{x}, \mathbf{p}, \kappa))$ , where  $\phi_{\lambda_1, \mathbf{p}}$  is the dominant eigenfunction of (17) for a fixed value  $\mathbf{p}$ . Note, however, that properties of  $\tilde{r}$  with respect to changes in parameters are harder to ascertain. This is because the Koopman eigenfunctions and eigenvalues change under parameter variations. Nevertheless the order is preserved in terms of the convergence times to a specific set  $\mathcal{A}_{\text{target}}$ .

Given the variations in the equilibria, eigenfunctions and eigenvalues, it seems almost impossible to estimate the convergence time to an isostable of the system (17) for a specific parameter value  $\mathbf{p}$  lying in the order-interval  $\mathbf{p} \in [\mathbf{p}_1, \mathbf{p}_2]$ . Interestingly, some estimates can still be provided. The result is quite technical and therefore, in what follows, we only sketch the main idea. Let us first introduce a few notations:

$$\begin{aligned}\mathcal{T}(\mathbf{x}, \kappa, \tau, \mathbf{p}, \varepsilon) &= \frac{1}{|\lambda_1(\mathbf{p})|} \ln \left( \frac{|\tilde{r}(\mathbf{x}, \kappa, \tau, \mathbf{p})|}{\varepsilon} \right), \\ \widetilde{\mathcal{T}}_\sigma(\mathbf{x}, \mathbf{p}, \varepsilon) &= \{(\kappa, \tau) \in \mathbb{R}_{\geq 0}^2 \mid \mathcal{T}(\mathbf{x}, \kappa, \tau, \mathbf{p}, \varepsilon) = \sigma\}, \\ \partial_- \mathcal{B}_\varepsilon(\mathbf{x}^*(\mathbf{p})) &= \{\mathbf{x} \in \mathbb{R}^n \mid \phi_{\lambda_1, \mathbf{p}} = -\varepsilon\} \text{ for } \varepsilon > 0, \\ \mathcal{S}_\sigma(\mathbf{x}, \mathbf{p}_1, \mathbf{p}_2, \varepsilon) &= \{(\kappa, \tau) \in \mathbb{R}^2 \mid v_1 \leq \mu \leq v_2, \xi_1 \leq \tau \leq \xi_2 \\ &\quad (v_1, \xi_1) \in \widetilde{\mathcal{T}}_\sigma(\mathbf{x}, \mathbf{p}_1, \varepsilon), (v_2, \xi_2) \in \widetilde{\mathcal{T}}_\sigma(\mathbf{x}, \mathbf{p}_2, \varepsilon)\}.\end{aligned}$$

It was shown in [29], that for any  $(\kappa, \tau) \in \mathcal{S}_\sigma(\mathbf{x}, \mathbf{p}_1, \mathbf{p}_2, \varepsilon)$  and  $\mathbf{p} \in [\mathbf{p}_1, \mathbf{p}_2]$ , the flow  $\mathbf{S}(t, \mathbf{x}, \mathbf{p}, \mu h(\cdot, \tau))$  at time  $\sigma + \tau$  belongs to the set  $\mathcal{A}_{\text{target}} = \{\mathbf{x} \in \mathbb{R}^n \mid \mathbf{z}_1 \preceq \mathbf{x} \preceq \mathbf{z}_2, \mathbf{z}_1 \in \partial_- \mathcal{B}_\varepsilon(\mathbf{x}^*(\mathbf{p}_1)), \mathbf{z}_2 \in \partial_- \mathcal{B}_\varepsilon(\mathbf{x}^*(\mathbf{p}_2))\}$ . This means that the set  $\widetilde{\mathcal{T}}_\sigma(\mathbf{x}, \mathbf{p}, \varepsilon)$  will intersect with  $\mathcal{S}_\sigma(\mathbf{x}, \mathbf{p}_1, \mathbf{p}_2, \varepsilon)$ , and for a small enough  $\varepsilon$  the set  $\mathcal{S}_\sigma(\mathbf{x}, \mathbf{p}_1, \mathbf{p}_2, \varepsilon)$  provides a good estimate on  $\widetilde{\mathcal{T}}_\sigma(\mathbf{x}, \mathbf{p}, \varepsilon)$  as demonstrated on numerical examples in [29].

## 5.2 Generalizing the Input Space

As we mentioned above the restrictions to single-input systems and temporal pulses can be relaxed. We will provide a generalization of the function  $r$  to control signals dependent on a vector of parameters as follows:

$$\mathbf{u}(t, \mathbf{p}, \tau) = \begin{cases} \mathbf{v}(t, \mathbf{p}) & \text{if } t \leq \tau \\ 0 & \text{otherwise} \end{cases} \quad (19)$$

where  $\mathbf{p} \in \mathcal{P} \subset \mathbb{R}_{\geq 0}^k$  is a parameter vector,  $\mathcal{P}$  is a convex, compact set, and  $\mathbf{v}(t, \mathbf{p}) : \mathbb{R} \times \mathbb{R}_{\geq 0}^k \rightarrow \mathbb{R}_{\geq 0}^m$  is an *a priori* given family of time dependent measurable functions. The dependence of  $\mathbf{v}$  on  $\mathbf{p}$  should be such that for any  $\mathbf{p}_1$  larger than  $\mathbf{p}_2$  in the standard partial order we have that  $\mathbf{u}(\cdot, \mathbf{p}_1)$  is larger than  $\mathbf{u}(\cdot, \mathbf{p}_2)$  in the standard partial order. For example, we can define a family of basis functions  $\{\mathbf{v}^i(t)\}_{i=1}^k : \mathbb{R} \rightarrow \mathbb{R}_{\geq 0}^m$  and parametrize control signals as

$$\mathbf{v}(t, \mathbf{p}) = \sum_{i=1}^k p_i \mathbf{v}^i(t).$$

Now we can define an extension of a pulse control function as

$$r(\mathbf{x}, \mathbf{p}, \tau) = \phi_{\lambda_1}(\mathbf{S}(\tau, \mathbf{x}, \mathbf{u}(\cdot, \mathbf{p}, \tau))).$$

This function can be computed in a similar manner to a pulsed control function  $r$  and it can also be verified that  $r(\mathbf{x}, \mathbf{p}, \tau)$  is a real function provided that the dominant eigenvalue  $\lambda_1$  is real.

Another parametrization can be useful in the multi-input case. We can define  $v_i$ , the  $i$ -th entry of  $\mathbf{v}$ , as  $v_i = \kappa_i h_i(t, p_i)$  and  $\tau = \max\{p_i\}$ . In this case, we can reproduce in a straightforward manner the results of Proposition 2, 3 and derive an optimization program similar to (14). Optimality can be obtained with additional regularity assumptions such as, for example, strong monotonicity for every  $\mathbf{p} \in \mathcal{P}$ .

### 5.3 Possible Relaxations of the Monotonicity Assumption

In order to use our results we need to make sure that the conclusion of Lemma 1 holds, that is  $\partial_\mu r(\mathbf{x}, \mu, \tau)$ ,  $\partial_{\mathbf{x}} r(\mathbf{x}, \mu, \tau)$  have nonnegative entries and  $\partial_\tau r(\mathbf{x}, \mu, \tau) > \lambda_1 r(\mathbf{x}, \mu, \tau)$ . It is fairly straightforward to show that we can relax monotonicity with respect to  $\mathbb{R}_{\geq 0}^n \times \mathbb{R}_{> 0}$  to monotonicity with respect to  $\mathcal{K} \times \mathbb{R}_{> 0}$ , where  $\mathcal{K}$  is a proper cone. Furthermore, the derivation of the optimization program (14) does not explicitly rely on the monotonicity assumption, except for the implicit use of Propositions 3 and 4. The key property of monotone systems allowing a “simple” optimal solution to (14) is the fact that the temporal pulses  $\kappa h(t, \tau)$  act in the direction of the gradient of  $\phi_{\lambda_1}(\mathbf{x})$ . Therefore if  $\phi_{\lambda_1}(\mathbf{x})$  is negative then temporal pulses are aligned with the flow of the unforced system, and if  $\phi_{\lambda_1}(\mathbf{x})$  is positive then temporal pulses act in the opposite direction. There is no indication that monotonicity is necessary and sufficient for this property. A particularly appealing relaxation for monotonicity is *differential positivity* [7], which was studied from the Koopman operator point of view in [18]. Differentially positive systems are defined for systems with a sufficiently smooth vector field  $F$  using a *prolonged* dynamical system:

$$\begin{aligned} \dot{\mathbf{x}} &= \mathbf{F}(\mathbf{x}, u), \\ \delta \dot{\mathbf{x}} &= \partial_{\mathbf{x}} \mathbf{F}(\mathbf{x}, u) \delta \mathbf{x} + \partial_u \mathbf{F}(\mathbf{x}, u) \delta u, \end{aligned} \quad (20)$$

with  $(\mathbf{x}, \delta \mathbf{x}) \in \mathbb{R}^n \times \mathbb{R}^n$  and where  $\partial_{\mathbf{x}}$ ,  $\partial_u$  denote the differentials with respect to  $\mathbf{x}$  and  $u$ , respectively. Following the definitions by [7], we let a *smooth cone field*  $\mathcal{K}(\mathbf{x})$  be defined as

$$\mathcal{K}(\mathbf{x}) = \left\{ \delta \mathbf{x} \in \mathbb{R}^n \mid k_i(\mathbf{x}, \delta \mathbf{x}) \geq 0, i = 1, \dots, m \right\},$$

where  $\mathcal{K}(\mathbf{x})$  is a proper cone for every  $\mathbf{x} \in \mathbb{R}^n$ , and  $k_i(\cdot, \cdot)$  are smooth functions. We will also use the dual cone field defined as

$$\mathcal{K}^*(\mathbf{x}) = \left\{ \mathbf{y} \in \mathbb{R}^n \mid \mathbf{y}^T \delta \mathbf{x} \geq 0, \delta \mathbf{x} \in \mathcal{K}(\mathbf{x}) \right\}.$$

Using the concepts above, the class of differentially positive systems is defined in the following way:

**Definition 4** *The system  $\dot{\mathbf{x}} = \mathbf{F}(\mathbf{x}, u)$  with a sufficiently smooth  $\mathbf{F}$  is called differentially positive with respect to the cone field  $\mathcal{K}_{\mathbf{x}}(\mathbf{x}, u) \times \mathcal{K}_u(\mathbf{x}, u)$  if the prolonged system leaves the cone field  $\mathcal{K}_{\mathbf{x}}(\mathbf{x}, u) \times \mathcal{K}_u(\mathbf{x}, u)$  invariant. Namely,*

$$\left. \begin{array}{l} \delta \mathbf{x}(t_0) \in \mathcal{K}_{\mathbf{x}}(\mathbf{x}(t_0), u(t_0)) \\ \delta u(t) \in \mathcal{K}_u(\mathbf{x}(t), u(t)) \forall t \geq t_0 \end{array} \right\} \Rightarrow \delta \mathbf{x}(t) \in \mathcal{K}_{\mathbf{x}}(\mathbf{x}(t), u(t)) \forall t \geq t_0. \quad (21)$$

In the case of dynamical systems  $\dot{\mathbf{x}} = \mathbf{F}(\mathbf{x})$  the prolonged system induces a flow  $(\mathbf{S}, \partial_{\mathbf{x}}\mathbf{S})$  such that  $(t, \mathbf{x}, \delta \mathbf{x}) \mapsto (\mathbf{S}(t, \mathbf{x}), \partial_{\mathbf{x}}\mathbf{S}(t, \mathbf{x})\delta \mathbf{x})$  is a solution of (20). Furthermore, if an exponentially stable equilibrium  $\mathbf{x}^*$  is such that the dominant eigenvalue  $\lambda_1$  of the Jacobian  $\mathbf{J}(\mathbf{x}^*)$  is simple and real then the system is differentially positive on the basin of attraction  $\mathcal{B}(\mathbf{x}^*)$ . Additionally, the cone field  $\mathcal{K}(\mathbf{x})$  can be built (under diagonalizability assumption on  $\mathbf{J}(\mathbf{x}^*)$ ) in such a way that the corresponding eigenfunction  $\phi_{\lambda_1}(\mathbf{x})$  lies in  $\mathcal{K}^*(\mathbf{x})$  for all  $\mathbf{x}$ . These properties seem to indicate that our results such as Proposition 4 and Lemma 1 can potentially be extended to differentially positive systems under additional restrictions. For instance, the case, where the cone field  $\mathcal{K}_{\mathbf{x}}(\mathbf{x}, u)$  does not depend on  $u$  ( $\mathcal{K}_{\mathbf{x}}(\mathbf{x}, u) = \mathcal{K}_{\mathbf{x}}(\mathbf{x})$ ) and  $\mathcal{K}_u(\mathbf{x}, u)$  is a positive orthant (i.e.,  $\mathcal{K}_u(\mathbf{x}, u) = \mathbb{R}_{\geq 0}^m$ ), appears as the simplest relaxation of monotonicity. However, differentially positive systems depending on control signals (i.e.,  $\dot{\mathbf{x}} = \mathbf{F}(\mathbf{x}, u)$ ) are not well-studied and many properties relevant to our problems are still not established. Furthermore, certificates for such systems would potentially involve checking conditions on the cone field for all  $\mathbf{x}$  in the state space, which appears to be computationally expensive. Therefore, the computational burden will be shifted from computing the optimal control to verifying system properties. On the other hand, if the solutions to (14) exist, then we can still compute them while perhaps sacrificing optimality. In this case, the approach can be used as a heuristic with a justification based on the results obtained for monotone systems.

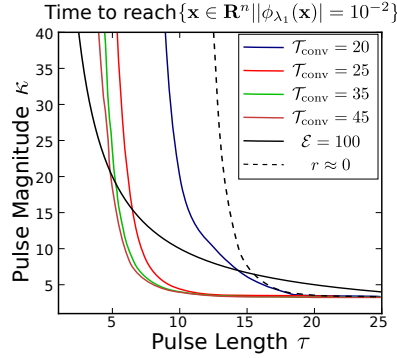
## 6 Numerical Examples

### 6.1 Controlling the Generalized Repressilator

The generalized repressilator represents an interesting case study analyzed theoretically in [37]. We consider a generalized repressilator with eight chemical species (states), which interact only with their direct neighbors in a ring topology. The corresponding dynamic equations are as follows:

$$\begin{aligned} \dot{x}_1 &= \frac{p_1}{1 + (x_8/p_2)^{p_3}} + p_4 - p_5 x_1 + u_1, \\ \dot{x}_2 &= \frac{p_1}{1 + (x_1/p_2)^{p_3}} + p_4 - p_5 x_2 + u_2, \\ \dot{x}_i &= \frac{p_1}{1 + (x_{i-1}/p_2)^{p_3}} + p_4 - p_5 x_i, \quad \forall i = 3, \dots, 8, \end{aligned} \quad (22)$$

where  $p_1 = 40$ ,  $p_2 = 1$ ,  $p_3 = 2$ ,  $p_4 = 1$ , and  $p_5 = 1$ . This system has two exponentially stable equilibria  $\mathbf{x}^*$  and  $\mathbf{x}^\bullet$  and is monotone with respect to the cones  $\mathcal{K}_{\mathbf{x}} = \mathbf{P}_{\mathbf{x}}\mathbb{R}^8$  and  $\mathcal{K}_{\mathbf{u}} = \mathbf{P}_{\mathbf{u}}\mathbb{R}^2$ , where  $\mathbf{P}_{\mathbf{x}} = \text{diag}([1, -1, 1, -1, 1, -1, 1, -1])$ ,



**Fig. 2** Level sets of  $\mathcal{T}_{\text{conv}}$  with  $\varepsilon = 10^{-2}$ , level set  $r = 0$ , and energy budget curve  $\kappa\tau = \mathcal{E}_{\text{max}}$ .

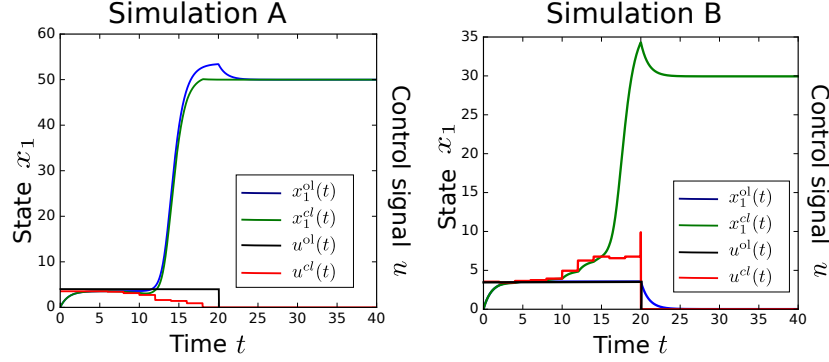
$\mathbf{P}_u = \text{diag}([1, -1])$ , and  $\mathbf{x}^\bullet \preceq_{\mathcal{K}_x} \mathbf{x}^*$ . It can actually be shown that the unforced system is strongly monotone in the interior of  $\mathbb{R}_{\geq 0}^8$  for all positive parameter values. The control signal  $u_1$  can switch the system from the state  $\mathbf{x}^\bullet$  to the state  $\mathbf{x}^*$ , while the control signal  $u_2$  can switch the system from the state  $\mathbf{x}^*$  to the state  $\mathbf{x}^\bullet$ .

### 6.1.1 Switching Problem

We first consider the problem of switching from  $\mathbf{x}^\bullet$  to  $\mathbf{x}^*$ . We present the comparison from [32] between the closed loop solution, which was discussed in this chapter, and the open loop solutions from [33]. The open loop solution computes the level sets of  $r(\mathbf{x}^\bullet, \kappa, \tau)$ , which allows to pick  $(\kappa, \tau)$  with an appropriate convergence time and spent energy. In terms of analysis it is more convenient to plot the level sets of the function  $\mathcal{T}_{\text{conv}}(\mathbf{x}^\bullet, \kappa, \tau, 10^{-2}) = \frac{1}{|\lambda_1|} \ln \left( \frac{r(\mathbf{x}^\bullet, \kappa, \tau)}{10^{-2}} \right) + \tau$  instead of the function  $r(\mathbf{x}^\bullet, \kappa, \tau)$ . Note that the function  $r$  is computed with the dominant eigenfunction associated with the target equilibrium  $\mathbf{x}^*$ . In Figure 2, we plot the level set of the function  $\mathcal{T}_{\text{conv}}(\mathbf{x}^\bullet, \kappa, \tau, 10^{-2}) = \frac{1}{|\lambda_1|} \ln \left( \frac{r(\mathbf{x}^\bullet, \kappa, \tau)}{10^{-2}} \right) + \tau$ , the level set  $\mathcal{E}_{\text{max}} = 100$  of the function  $\kappa\tau = \mathcal{E}_{\text{max}}$ , and the level set  $r(\mathbf{x}^\bullet, \kappa, \tau) = 0$ . The last two are related to the constraints of the static optimization program (14).

The function  $\mathcal{T}_{\text{conv}}$  can escape to  $-\infty$  around the level set  $r(\mathbf{x}^\bullet, \kappa, \tau) \approx 0$ . This is not a conflict with the interpretation of the function  $\mathcal{T}_{\text{conv}}$ , since it represents the convergence time only if the value of  $|r(\mathbf{x}^\bullet, \kappa, \tau)|$  is larger than  $10^{-2}$ . Otherwise  $\frac{1}{|\lambda_1|} \ln \left( \frac{|r(\mathbf{x}^\bullet, \kappa, \tau)|}{10^{-2}} \right)$  is negative, and the computational results are meaningless. This also explains why the level sets of  $\mathcal{T}_{\text{conv}}$  appear to have the same asymptotics as the level set  $r(\mathbf{x}^\bullet, \kappa, \tau) = 0$  in Figure 2.

The closed-loop approach relies on computing the function  $r(\mathbf{x}, \kappa, \tau)$  at a given point  $\mathbf{x}$ , which is not necessarily an equilibrium. While it is fairly clear that the closed-loop solutions are preferable, we still compare the closed and open loop



**Fig. 3** Closed-loop and open-loop switching. Left and right panels depict simulations for Settings A and B, respectively. In both figures,  $x_1^{ol}$ ,  $x_1^{cl}$  stand for the trajectories of the state  $x_1$  in the open and closed-loop settings, respectively, and  $u^{ol}$ ,  $u^{cl}$  stand for the corresponding control signals.

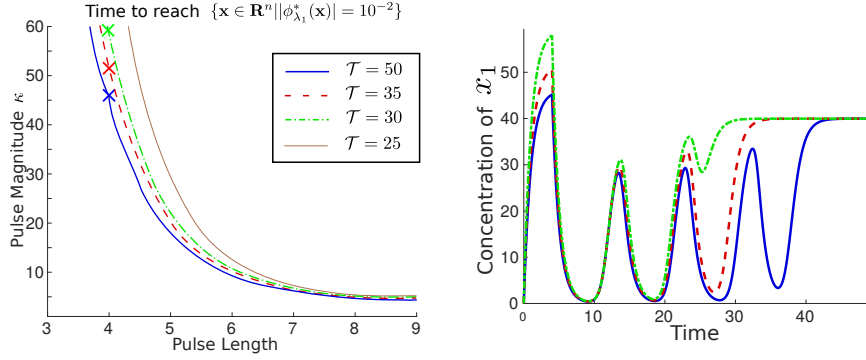
strategies subject to perturbations of parameters  $p_{ij}$ . In particular we compute the control signals using nominal parameter values, but the simulations are performed using modified parameters

*Setting A.* We set  $p_{i1}^A = 50$  for odd  $i$ .

*Setting B.* We set  $p_{i1}^B = 30$  for odd  $i$ .

The Euclidean distance between the nominal initial point and the actual initial point in Settings A and B is equal to 0.025 and 0.031, respectively, which is negligible.

In order to compute an open-loop optimal control policy based on the nominal model (i.e. with parameter values  $p_{ij}$ ), one can solve the static optimization program (14). Figure 2 also offers a graphical solution to the problem and a depiction of possible tradeoffs in the problem. In our case, the optimal solution lies at the intersection of the constraint curves (i.e. energy budget curve and level set  $r(\mathbf{x}^*, \kappa, \tau) = 0$ ). We set  $\mathcal{E}_{\max} = 100$  and for both closed and open-loop strategies we pick a pair  $(\kappa^0, \tau^0) = (3.53, 20)$  lying near the zero level set of  $r$ , which is not on the energy budget constraint curve. We make this choice, in order to be able to react to possible disturbances/modeling errors. The closed-loop control is updated every  $t_{\text{samp}} = 2$  time units. At  $i$ -th iteration the energy budget is updated for the energy spent. Then the values of  $r(\mathbf{x}((i-1)t_{\text{samp}}), \kappa, \tau - (i-1)t_{\text{samp}})$  are computed, where  $\mathbf{x}((i-1)t_{\text{samp}})$  is the current state of the system, and the values of  $\kappa$  are taken on a uniform grid of 100 points in  $[2, 10]$ . Thereafter  $\kappa^0$  is updated by minimizing the convergence time  $\mathcal{T}_{\text{conv}}$  using computed values of  $r$  satisfying the terminal set and the energy budget constraint. The simulation results are depicted in Figure 3. In Simulation A, the system converges to the target equilibrium faster than the nominal one (i.e. with parameters  $p_{ij}$ ) and the closed-loop solution saves energy and limits the overshoot in comparison with the open-loop solution. In Simulation B, the system converges to the target equilibrium slower than the nominal one, and all the energy budget is spent. In this case, the closed-loop solution allows the switch, while the open-loop (i.e. [33]) does not.



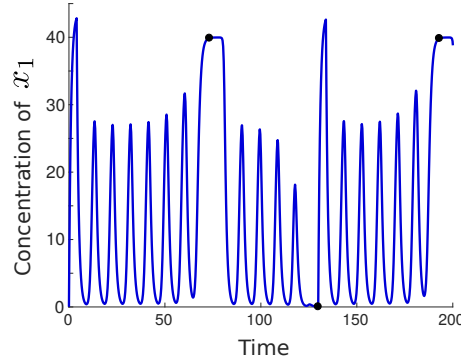
**Fig. 4** Switching between equilibria in a generalized repressilator system. In the left panel, the level sets of the curve  $1/\Re(\lambda_1) \ln(\phi_{\lambda_1}^*(\mathbf{x})/10^{-2})$  are depicted. The trajectories in the right panel are generated by the pairs  $(\kappa, \tau)$  corresponding to the crosses on the left panel. Furthermore, for any pair  $(\kappa, \tau)$  lying on green dash-dotted, red dashed, blue solid level set of  $1/\Re(\lambda_1) \ln(\phi_{\lambda_1}^*(\mathbf{x})/10^{-2})$  (the left panel) the corresponding trajectory will have a similar convergence time  $\mathcal{T}$  as green dash-dotted, red dashed, blue solid trajectories (the right panel), respectively.

### 6.1.2 “Stabilizing” an Unstable Periodic Orbit in the Generalized Repressilator.

In [37] it was shown that the generalized repressilator admits an unstable periodic orbit, which lies around the saddle point located in the order-interval  $[\mathbf{x}^\bullet, \mathbf{x}^*]$ . Therefore, oscillations can be induced by pushing the trajectories from one stable equilibrium to another repeatedly. Here we also rely on pulse control functions, but in this case we have two:  $r^*(\mathbf{x}^\bullet, \kappa, \tau)$  (for converging to  $\mathbf{x}^*$ ) and  $r^\bullet$  (for converging to  $\mathbf{x}^\bullet$ ) defined using the eigenfunctions  $\phi_{\lambda_1}^*(\mathbf{x})$  and  $\phi_{\lambda_1}^\bullet(\mathbf{x})$ , respectively, which in turn are defined on the basins of attraction  $\mathcal{B}(\mathbf{x}^*)$  and  $\mathcal{B}(\mathbf{x}^\bullet)$ , respectively. Since the parameters in every  $i$ -th equation are the same we can compute only the function  $r^*$  and deduce  $r^\bullet$  by renumbering the states.

In the left panel Figure 4 we plot the level sets of the function  $\mathcal{T} = \ln\left(\frac{|r^*(\mathbf{x}^\bullet, \kappa, \tau)|}{10^{-2}}\right)$ , where crosses stand for representative pairs  $(\kappa, \tau)$  and the trajectories corresponding to those crosses are plotted in right panel of Figure 4. As the reader may notice, once the function  $r^*$  is computed the choice of the control signal is rather straightforward. As demonstrated in [32], the pulse control function can also be used for closed-loop switching where the pair  $(\kappa, \tau)$  is updated at prescribed time intervals. Here, we will use the pulse control functions in order to induce oscillations. Numerical simulations suggest that the trajectories follow a quasi-periodic orbit in Figure 4, while switching between the stable equilibria using a pulse. Furthermore, the oscillations can be sustained for an arbitrary long time, provided that the flow  $\mathbf{S}(t, x_0, \kappa h(\cdot, \tau))$  at time  $\tau$  lies on the manifold separating the basins of attraction  $\mathcal{B}(\mathbf{x}^*)$  and  $\mathcal{B}(\mathbf{x}^\bullet)$ . This orbit, however, is unstable since an arbitrary small deviation from this manifold will result in convergence of the flow to  $\mathbf{x}^*$  or  $\mathbf{x}^\bullet$ . For this problem we will force the flow to remain in a compact set  $\mathcal{M} = [\mathbf{x}^\bullet + \delta \mathbf{P}_x \mathbf{1}, \mathbf{x}^* - \delta \mathbf{P}_x \mathbf{1}]$ , where  $\delta = 1$  and  $\mathbf{P}_x$





**Fig. 5** Inducing oscillatory behavior in the generalized repressilator system with eight states. The pulses for both  $u_1$  and  $u_2$  are equal, and are generated using a pair  $(\kappa, \tau) = (44.5, 4)$ . The control signal  $u_1$  is applied at times  $t = 0, 111.92$ , while the control signal  $u_2$  is applied at times  $t = 57.97, 165.8$ .

is the matrix defining the partial order in the state variables. Note that the set  $\mathcal{M}$  contains the unstable periodic orbit for the chosen parameter values. We will update the control signal only when the flow leaves  $\mathcal{M}$ . For both  $u_1$ , and  $u_2$  we use the pair  $(\kappa, \tau) = (44.5, 4)$  to compute the control signals. The control signal  $u_1 = \kappa h(\cdot, \tau)$  (respectively,  $u_2 = \kappa h(\cdot, \tau)$ ) forces the switch from  $\mathbf{x}^\bullet$  to  $\mathbf{x}^*$  (respectively, from  $\mathbf{x}^*$  to  $\mathbf{x}^\bullet$ ), while resulting in a trajectory with lightly damped oscillations. The control signal is switched from  $u_1$  to  $u_2$ , when the flow leaves  $\mathcal{M}$  while approaching  $\mathbf{x}^*$ . Meaning that we set  $u_2 = \kappa h(\cdot, \tau)$  and  $u_1 = 0$ . The control is switched from  $u_2$  to  $u_1$ , when the flow leaves  $\mathcal{M}$  while approaching  $\mathbf{x}^\bullet$ . We depict the simulations in Figure 5.

More uniform oscillations may potentially be obtained if we keep the flow inside the set

$$\tilde{\mathcal{M}} = \{\mathbf{z} \in \mathbb{R}_{\geq 0}^8 \mid \phi_{\lambda_1}^*(\mathbf{z}) \geq 1/\delta \text{ and } \phi_{\lambda_1}^*(\mathbf{z}) \leq -1/\delta\},$$

for a small positive  $\delta$ . This can be done by applying the control signal  $u_1 = \kappa_1 h(\cdot, t_1 + \tau_1)$  if the flow reaches the set  $\{\mathbf{z} \in \mathbb{R}_{\geq 0}^8 \mid \phi_{\lambda_1}^*(\mathbf{z}) = 1/\delta\}$  at the time  $t_1$ , and the control signal  $u_2 = \kappa_2 h(\cdot, t_2 + \tau_2)$  if the flow reaches the set  $\{\mathbf{z} \in \mathbb{R}_{\geq 0}^8 \mid \phi_{\lambda_1}^*(\mathbf{z}) = -1/\delta\}$  at the time  $t_2$ . The values of  $(\kappa_1, \tau_1)$ ,  $(\kappa_2, \tau_2)$  are chosen using the functions  $r^*(\mathbf{x}(t_1), \cdot, \cdot)$ ,  $r^*(\mathbf{x}(t_2), \cdot, \cdot)$ , respectively, where  $\mathbf{x}(t)$  denotes the value of the state  $\mathbf{x}$  at the time  $t$ .

We note that a similar control strategy was used in [37]. However, no computational approach to estimate the settling time of the trajectory, the pulse magnitude or its length was provided. In [35], it was proposed to track other periodic trajectories (such as, sine functions of time) instead using approximate dynamic programming. However, the solution was very computationally expensive and offered little insight into the problem. We finally note that the existence of an unstable periodic orbit is not required to induce a periodic orbit in a monotone system as was demonstrated in

*vivo* on a toggle switch system implemented in *E. Coli* in [16]. Again the theoretical results in [29] justified an intuitive approach in [16].

## 6.2 HIV Viral Load Control

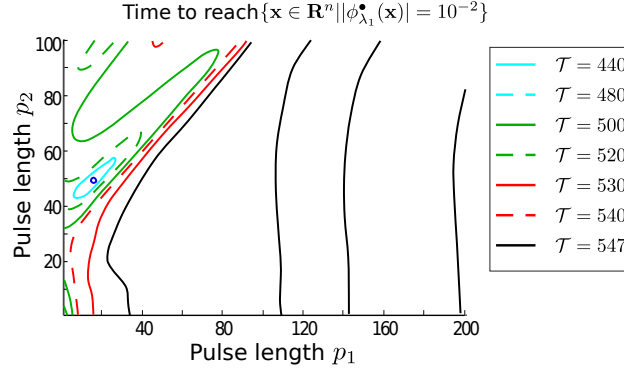
Viral load dynamics of the human immunodeficiency virus (HIV) in the bloodstream can be modeled by a bistable system, where one stable equilibrium corresponds to the high viral load (non-healthy state) and the other to the low viral load (healthy state). Even though HIV is still present in “the healthy state”, the virus population is suppressed by the immune system and cannot increase to life-threatening levels. Drug therapies can lower the viral load, once a therapy is interrupted, however, the viral load can return to the non-healthy state. Mathematically this means that the trajectory of the model does not reach the basin of attraction of the healthy state under this therapy protocol. Typically, a combination of therapies is used, which mathematically can be expressed as temporal control signals. The objective is to design the therapies (control signals) so as to drive the viral load to the “healthy state”. The therapies typically follow structured treatment interruption (STI) protocols, that is, the therapies are interrupted at specific times. In this example, we will validate this approach based on mathematical modelling. We take the modeling approach developed in [1] and consider the following differential equations:

$$\begin{aligned}\dot{T}_1 &= s_1 - \gamma_1 T_1 - (1 - u_1)\beta_1 V T_1 \\ \dot{T}_2 &= s_2 - \gamma_2 T_2 - (1 - f u_1)\beta_2 V T_2 \\ \dot{T}_1^* &= (1 - u_1)\beta_1 V T_1 - c_{T^*} T_1^* - \kappa_1 E T_1^* \\ \dot{T}_2^* &= (1 - f u_1)\beta_2 V T_2 - c_{T^*} T_2^* - \kappa_2 E T_2^* \\ \dot{V} &= (1 - u_2)N_T k(T_1^* + T_2^*) - ((1 - u_1)\rho_1 \beta_1 T_1 + (1 - f u_1)\rho_2 \beta_2 T_2)V - c_V V \\ \dot{E} &= \lambda_E + \frac{b_E(T_1^* + T_2^*)}{(T_1^* + T_2^*) + K_b} E - \frac{d_E(T_1^* + T_2^*)}{(T_1^* + T_2^*) + K_d} E - c_E E,\end{aligned}$$

where  $T_1$  is the number of healthy  $CD4^+$  T-lymphocytes,  $T_2$  is the number of healthy macrophages,  $T_1^*$  is the number of infected  $CD4^+$  T-lymphocytes,  $T_2^*$  is the number of infected macrophages,  $V$  is the number of free virus particles and  $E$  is the number of HIV-specific cytotoxic T-cells. Given the parameters

$$\begin{aligned}s_1 &= 10^4, \gamma_1 = \gamma_2 = 0.01, \beta_1 = 8 \cdot 10^{-7}, s_2 = 31.98, \\ f &= 0.34, \beta_2 = 10^{-4}, c_{T^*} = 0.7, \kappa_1 = \kappa_2 = 10^{-5}, c_V = 13 \\ N_T &= K_b = 100, \rho_1 = \rho_2 = \lambda_E = 1, k = 0.7, b_E = 0.3, \\ d_E &= 0.25, K_d = 500, c_E = 0.1,\end{aligned}$$

the system is not monotone and it has three equilibria, which will denote  $\mathbf{x}^u$ ,  $\mathbf{x}^*$  and  $\mathbf{x}^\bullet$ . Computation with the Matlab symbolic toolbox gives the following values:



**Fig. 6** Level sets of the convergence time to the equilibrium  $\mathbf{x}^*$  (non-healthy state) for different times  $p_1, p_2$ . The convergence time is a continuous function (as it is a function of the  $C^1$  dominant eigenfunction), and hence for the considered combinations of  $p_1$  and  $p_2$  the switch to the healthy state cannot be performed. Furthermore, numerical simulations suggest that the convergence time is bounded from above and hence the flow lies relatively far from the manifold separating the basins of attraction. This indicates that continuous therapies are not effective as an HIV treatment protocol.

$$\begin{aligned}\mathbf{x}^u &= (6.64 \cdot 10^5, 49.97, 1.21 \cdot 10^3, 11.33, 6.29 \cdot 10^3, 2.07 \cdot 10^5), \\ \mathbf{x}^\bullet &= (1.63 \cdot 10^5, 4.99, 1.19 \cdot 10^4, 45.59, 6.39 \cdot 10^4, 23.54), \\ \mathbf{x}^* &= (9.67 \cdot 10^5, 620.51, 76.01, 6.09, 415.37, 3.5311 \cdot 10^5).\end{aligned}$$

The equilibrium  $\mathbf{x}^u$  is a saddle with an unstable manifold of dimension 1;  $\mathbf{x}^*$  corresponds to the “healthy” state, and  $\mathbf{x}^\bullet$  corresponds to the “non-healthy” state. The equilibria  $\mathbf{x}^\bullet, \mathbf{x}^*$  are exponentially stable with complex dominant eigenvalues, which rules out monotonicity. In [1], the authors considered an optimal control problem, which they have addressed using an approximated solution to the Pontryagin’s maximum principle. The optimal therapies  $u_1$  and  $u_2$  take only values between zero and one, however, it is noticeable that these strategies are not pulses (Figure 4 in [1]). We will show that switching from the unhealthy state to the healthy one is highly unlikely using temporal pulses.

In [1, 6], the optimal therapies were designed in such a way that the maximal values of  $u_1$  and  $u_2$  are equal to 0.7 and 0.3, respectively. Therefore, we assume that

$$\begin{aligned}u_1(t) &= 0.7h(t, p_1) \\ u_2(t) &= 0.3h(t, p_2)\end{aligned}$$

This means that we have a vector  $\mathbf{p}$  with two entries and  $\tau = \max(p_1, p_2)$ . Therefore, we can use the discussion above in order to compute the function  $r(\mathbf{x}^\bullet, \mathbf{p}, \tau)$  defined using the eigenfunction  $\phi_{\lambda_1}$  around the equilibrium  $\mathbf{x}^\bullet$ . Since  $\tau$  is equal to one of the two  $p_i$ , we can plot the function  $r$  in the parameter space  $p_1, p_2$ . The computation results are depicted in Figure 6.

It is noticeable that the function  $r$  is finite for all  $\mathbf{p}$ , which means that the switch to  $\mathbf{x}^*$  is not occurring. Moreover, when  $p_2 \rightarrow \infty$ , the flow converges to the origin. As can be verified through simulation, small nonnegative state values of  $\mathbf{x}$  around 0 lie in the basin of attraction of  $\mathbf{x}^*$ . This means that the switching to  $\mathbf{x}^*$  using these types of pulses is not possible. This validates the STI approach studied in [1], according to which both therapies have to be interrupted at specific times in order to reach the “healthy” state  $\mathbf{x}^*$ . The more interesting and important question of how to design (not necessarily optimally) the control signals  $u_1$  and  $u_2$  is currently hard to answer using the presented approach.

We note that some bistable systems with complex dominant eigenvalues of the Jacobian at the equilibria are amenable to switching using pulses. In fact, the behavior of these systems can be extremely complex. In [33], it was shown that such a control strategy is successful for switching between equilibria in a Lorentz system, which has chaotic attractors.

## 7 Acknowledgement

Most of the work was conducted when Dr Sootla held F.R.S-FNRS fellowship at University of Liège. Currently, Dr Sootla is supported by UK’s Engineering and Physical Sciences (EPSRC) Grant EP/M002454/1. Dr Stan gratefully acknowledges support from the U.K. EPSRC through the EPSRC Fellowship EP/M002187/1.

## 8 Conclusion

In this chapter, we discussed how to use the Koopman operator (and its eigenfunctions) to solve the problem of convergence to an equilibrium. The problem is cast using the concept of isostables — level sets of the dominant Koopman eigenfunction. For monotone systems, this reformulation allows rewriting of the optimal control problem as a static optimization program under additional regularity assumptions. The core idea of the method is to use isostables as a target set for the optimal control problem, which simplifies the computation of time-parametrized optimal control signals. This method can be applied to control bistable systems, if the control variable, considered as a constant parameter, forces a bifurcation in the system, and the resulting equilibrium is attractive and lies in the basin of attraction of the target equilibrium. Therefore, the approach can be applied to non-monotone systems, at the cost of a potential loss of optimality. For example, this approach was applied to the synchronization of cardiac cells [32], event-based regulation around a saddle point [29], switching under parametric uncertainty [29], and the problem of inducing oscillations as discussed above. At the moment optimality of solutions is guaranteed for the class of *monotone* systems, but as we discussed in Section 5.3 optimality results can potentially be extended to a larger class of systems. Another

interesting question is whether Koopman eigenfunctions (or their properties) can be used to predict a parametric form of control signals that provide solutions to the convergence problem. Advances in the data-based Koopman operator algorithms such as Dynamic Mode Decomposition (DMD), can potentially open the door for data-efficient data-based control methods.

## References

1. B.M. Adams, H.T. Banks, H.-D. Kwon, and H. T. Tran. Dynamic multidrug therapies for HIV: Optimal and STI control approaches. *Math Biosci Eng*, 1(2):223–241, 2004.
2. D. Angeli and E.D. Sontag. Monotone control systems. *IEEE Trans Autom Control*, 48(10):1684–1698, 2003.
3. A. Berman and R.J. Plemmons. *Nonnegative Matrices in the Mathematical Sciences*, volume 9. SIAM, 1994.
4. J.A.N. Brophy and C.A. Voigt. Principles of genetic circuit design. *Nat methods*, 11(5):508–520, 2014.
5. D. Ernst, P. Geurts, and L. Wehenkel. Tree-based batch mode reinforcement learning. *J Mach Learn Res*, 6:503–556, 2005.
6. D. Ernst, G.-B. Stan, J. Goncalves, and L. Wehenkel. Clinical data based optimal STI strategies for HIV: a reinforcement learning approach. In *IEEE Conf Decision Control*, pages 667–672. IEEE, 2006.
7. F. Forni and R. Sepulchre. Differentially positive systems. *IEEE Trans Autom Control*, 61(2):346–359, 2016.
8. Paul S Freemont and Richard I Kitney. *Synthetic Biology-a Primer (revised Edition)*. World Scientific, 2015.
9. M.W. Hirsch, H. Smith, et al. Monotone dynamical systems. *Handbook of differential equations: ordinary differential equations*, 2:239–357, 2005.
10. V. Hsiao, A. Swaminathan, and R.M. Murray. Control theory for synthetic biology: Recent advances in system characterization, control design, and controller implementation for synthetic biology. *IEEE Control Systems*, 38(3):32–62, 2018.
11. Linda J Kahl and Drew Endy. A survey of enabling technologies in synthetic biology. *Journal of biological engineering*, 7(1):13, 2013.
12. E. Kaiser, J.N. Kutz, and S.L. Brunton. Data-driven discovery of koopman eigenfunctions for control. *arXiv preprint arXiv:1707.01146*, 2017.
13. E.S. Kim, M. Arcak, and S.A. Seshia. Directed specifications and assumption mining for monotone dynamical systems. In *Proc Conf Hybrid Systems: Computation Control*, pages 21–30. ACM, 2016.
14. Milan Korda and Igor Mezić. Linear predictors for nonlinear dynamical systems: Koopman operator meets model predictive control. *Automatica*, 93:149–160, 2018.
15. Y. Lan and I. Mezić. Linearization in the large of nonlinear systems and Koopman operator spectrum. *Physica D*, 242:42–53, 2013.
16. J.-B. Lugagne, S.S. Carrillo, M. Kirch, A. Köhler, G. Batt, and P. Hersen. Balancing a genetic toggle switch by real-time feedback control and periodic forcing. *Nature Communications*, 8(1):1671, 2017.
17. A. Mauroy. Converging to and escaping from the global equilibrium: Isostables and optimal control. In *IEEE Conf Decision Control*, pages 5888–5893, 2014.
18. A. Mauroy, F. Forni, and R. Sepulchre. An operator-theoretic approach to differential positivity. In *IEEE Conf Decision Control*, pages 7028–7033, Dec 2015.
19. A. Mauroy and I. Mezić. Global stability analysis using the eigenfunctions of the koopman operator. *IEEE Trans Autom Control*, 61(11):3356–3369, 2016.

20. A. Mauroy, I. Mezić, and J. Moehlis. Isostables, isochrons, and koopman spectrum for the action–angle representation of stable fixed point dynamics. *Physica D*, 261:19–30, 2013.
21. F. Menolascina, M. Di Bernardo, and D. Di Bernardo. Analysis, design and implementation of a novel scheme for in-vivo control of synthetic gene regulatory networks. *Automatica, Special Issue on Systems Biology*, 47(6):1265–1270, April 2011.
22. A. Miliadis-Argeitis, S. Summers, J. Stewart-Ornstein, I. Zuleta, D. Pincus, H. El-Samad, M. Khammash, and J. Lygeros. In silico feedback for in vivo regulation of a gene expression circuit. *Nat biotechnol*, 29(12):1114–1116, 2011.
23. S. Peitz. Controlling nonlinear PDEs using low-dimensional bilinear approximations obtained from data. *arXiv preprint arXiv:1801.06419*, 2018.
24. Sebastian Peitz and Stefan Klus. Koopman operator-based model reduction for switched-system control of PDEs. *arXiv preprint arXiv:1710.06759*, 2017.
25. P.E.M. Purnick and R. Weiss. The second wave of synthetic biology: from modules to systems. *Nat. Rev. Mol. Cell Biol.*, 10(6):410–422, Jun 2009.
26. P. J. Schmid. Dynamic mode decomposition of numerical and experimental data. *Journal of Fluid Mechanics*, 656:5–28, 2010.
27. E.D. Sontag. Monotone and near-monotone biochemical networks. *Syst Synthetic Biol*, 1(2):59–87, 2007.
28. E.D. Sontag. *Mathematical control theory: deterministic finite dimensional systems*, volume 6. Springer Science & Business Media, 2013.
29. A. Sootla and D. Ernst. Pulse-based control using Koopman operator under parametric uncertainty. *IEEE Tran Autom Control*, 63(3):791–796, March 2018.
30. A. Sootla and A. Mauroy. Properties of isostables and basins of attraction of monotone systems. In *Proc Am Control Conf*, pages 7365–7370, 2016.
31. A. Sootla and A. Mauroy. Geometric properties and computation of isostables and basins of attraction of monotone systems. *IEEE Trans Autom Control*, 62(12):6183–6194, 2017.
32. A. Sootla, A. Mauroy, and D. Ernst. An optimal control formulation of pulse-based control using Koopman operator. *Automatica*, 91:217 – 224, 2018.
33. A. Sootla, A. Mauroy, and J. Gonçalves. Shaping pulses to control monotone bistable systems using Koopman operator. In *Proc Symp Nonlin Control Systems*, pages 710–715, Aug 2016.
34. A. Sootla, D. Oyarzún, D. Angeli, and G.-B. Stan. Shaping pulses to control bistable systems analysis, computation and counterexamples. *Automatica*, 63:254–264, Jan. 2016.
35. A. Sootla, N. Strelkova, D. Ernst, M. Barahona, and G.-B. Stan. On reference tracking using reinforcement learning with application to gene regulatory networks. In *Proc Conf. Decision Control*, pages 4086–4091, Florence, Italy, Dec. 10-13 2013.
36. A. Sootla, N. Strelkova, D. Ernst, M. Barahona, and G.-B. Stan. Toggling the genetic switch using reinforcement learning. In *Proc French Meeting on Planning, Decision Making and Learning*, Liège, Belgium, May 2014.
37. N. Strelkova and M. Barahona. Switchable genetic oscillator operating in quasi-stable mode. *J R Soc Interface*, 7(48):1071–1082, 2010.
38. J. H. Tu, C. W. Rowley, D. M. Luchtenburg, S. L. Brunton, and J. N. Kutz. On dynamic mode decomposition: Theory and applications. *J Comput Dynamics*, 1(2):391 – 421, Dec 2014.
39. J. Uhlendorf, A. Miermont, T. Delaveau, G. Charvin, F. Fages, S. Bottani, G. Batt, and P. Hersen. Long-term model predictive control of gene expression at the population and single-cell levels. *Proc. Nat. Academy Sciences*, 109(35):14271–14276, 2012.
40. D. Wilson and J. Moehlis. An energy-optimal methodology for synchronization of excitable media. *SIAM J Applied Dynamical Systems*, 13(2):944–957, 2014.