Optimal trajectory tracking

vorgelegt von
Diplom Physiker
Jakob Löber
aus Erfurt

Von der Fakultät II - Mathematik und Naturwissenschaften
der Technischen Universität Berlin
zur Erlangung des akademischen Grades
Doktor der Naturwissenschaften

Dr. rer. nat.

genehmigte Dissertation

Promotionsausschuss:

Vorsitzender:	Prof. Dr. Dieter Breitschwerdt
Berichter:	Prof. Dr. Harald Engel
Berichter:	Prof. Dr. Alexander S. Mikhailov
Berichter:	Prof. Dr. Fredi Tröltzsch

Tag der wissenschaftlichen Aussprache: 6. Juli 2015

Berlin 2015

Optimal trajectory tracking

Jakob Löber

Summary

This thesis investigates optimal trajectory tracking of nonlinear dynamical systems with affine controls. The control task is to enforce the system state to follow a prescribed desired trajectory as closely as possible. The concept of so-called exactly realizable trajectories is proposed. For exactly realizable desired trajectories exists a control signal which enforces the state to exactly follow the desired trajectory.
This approach does not only yield an explicit expression for the control signal in terms of the desired trajectory, but also identifies a particularly simple class of nonlinear control systems. Systems in this class satisfy the so-called linearizing assumption and share many properties with linear control systems. For example, conditions for controllability can be formulated in terms of a rank condition for a controllability matrix analogously to the Kalman rank condition for linear time invariant systems.
Furthermore, exactly realizable trajectories arise as solutions to unregularized optimal control problems. Based on that insight, the regularization parameter is used as the small parameter for a perturbation expansion. This results in a reinterpretation of affine optimal control problems with small regularization term as singularly perturbed differential equations. The small parameter originates from the formulation of the control problem and does not involve simplifying assumptions about the system dynamics. Combining this approach with the linearizing assumption, approximate and partly linear equations for the optimal trajectory tracking of arbitrary desired trajectories are derived.
For vanishing regularization parameter, the state trajectory becomes discontinuous and the control signal diverges. On the other hand, the analytical treatment becomes exact and the solutions are exclusively governed by linear differential equations. Thus, the possibility of linear structures underlying nonlinear optimal control is revealed. This fact enables the derivation of exact analytical solutions to an entire class of nonlinear trajectory tracking problems with affine controls. This class comprises, among others, mechanical control systems in one spatial dimension and the FitzHugh-Nagumo model with a control acting on the activator.

Zusammenfassung

Die vorliegende Arbeit behandelt die optimale Bahnverfolgung in nichtlinearen dynamischen Systemen mit linear eingehender Kontrolle. Das Ziel der Kontrolle ist es, den Systemzustand so nah wie möglich entlang einer vorgeschriebenen Referenztrajektorie zu steuern. Das Konzept der sogenannten exakt realisierbaren Referenztrajektorie wird eingeführt. Für exakt realisierbare Referenztrajektorien existieren Kontrollsignale, welche den Zustand exakt entlang der gewünschten Trajektorie steuern.
Dieser Zugang ergibt nicht nur einen Ausdruck für das Kontrollsignal als Funktion der Referenztrajektorie, sondern identifiziert auch eine besonders einfache Klasse von nichtlinearen Kontrollsystemen. Diese Klasse erfüllt die sogenannte Linearisierungsvorraussetzung und teilt viele Eigenschaften mit linearen Kontrollsystemen. Zum Beispiel kann die Kontrollierbarkeit dieser Klasse mithilfe einer Rangbedingung für eine Kontrollierbarkeitsmatrix analog zur Kalman'schen Rangbedingung für lineare zeitinvariante Systeme formuliert werden.
Darüberhinaus ergeben sich exakt realisierbare Referenztrajektorien als Lösung unregularisierter optimaler Kontrollprobleme. Aufbauend auf diesem Resultat wird der Regularisierungsparameter als kleiner Parameter einer Störungsentwicklung benutzt. Dies führt zu einer Neuinterpretation optimaler Kontrollprobleme mit kleinem Regularisierungsparameter und linear eingehender Kontrolle als Systeme singulär gestörter Differentialgleichungen. Der kleine Parameter resultiert einzig und allein aus der Formulierung des Kontrollproblems und benötigt keine vereinfachenden Annahmen über die Systemdynamik. Kombiniert man diesen Ansatz mit der Linearisierungsvoraussetzung, ergeben sich teilweise lineare Näherungsgleichungen für die optimale Bahnverfolgung beliebiger Referenztrajektorien.
Für einen verschwindenden Regularisierungsparameter ist der Zustand eine unstetige Funktion der Zeit und die Kontrolle divergiert. Andererseits jedoch sind die abgeleiteten Gleichungen exakt und ausschließlich linear. Hiermit wird die Möglichkeit zugrundeliegender linearer Strukturen in nichtlinearer optimaler Kontrolle aufgezeigt. Dies ermöglicht die Ableitung exakter analytischer Lösungen für eine ganze Klasse nichtlinearer Bahnverfolgungsprobleme mit linear eingehender Kontrolle. Diese Klasse umfasst unter anderem mechanische Kontrollsysteme in einer räumlichen Dimension sowie das FitzHugh-Nagumo Modell mit einer auf den Aktivator wirkenden Kontrolle.

Contents

List of Figures

1. Introduction

Science often begins with the discovery of physical phenomena. The second step is to describe, understand, and predict them, often in terms of mathematical theories. The final step is to take advantage of the discovered phenomena. This last step is the topic of control theory.

Section 1.1 introduces the notation for control systems. Some examples of affine control systems, which are used repeatedly throughout the thesis to demonstrate the developed concepts, are presented in Section 1.2. Section 1.3 illustrates the main result of this thesis by means of an example.

1.1. Affine control systems

The subject of this thesis are controlled dynamical systems of the form

$$\dot{\boldsymbol{x}}(t) = \boldsymbol{R}(\boldsymbol{x}(t)) + \boldsymbol{\mathcal{B}}(\boldsymbol{x}(t))\,\boldsymbol{u}(t), \tag{1.1}$$

$$\boldsymbol{x}(t_0) = \boldsymbol{x}_0. \tag{1.2}$$

Here, t is the time, $\boldsymbol{x}(t) = (x_1(t), \ldots, x_n(t))^T \in \mathbb{R}^n$ is called the *state vector* with n components and $\boldsymbol{x}^T$ denotes the transposed of vector $\boldsymbol{x}$. The dot

$$\dot{\boldsymbol{x}}(t) = \frac{d}{dt}\boldsymbol{x}(t) \tag{1.3}$$

denotes the time derivative of $\boldsymbol{x}(t)$. The vector $\boldsymbol{u}(t) = (u_1(t), \ldots, u_p(t))^T \in \mathbb{R}^p$ with $p \leq n$ components is the vector of *control* or *input signals*. The *nonlinearity* $\boldsymbol{R}$ is a sufficiently well behaved function mapping $\mathbb{R}^n$ to $\mathbb{R}^n$, and $\boldsymbol{\mathcal{B}}$ is a sufficiently well behaved $n \times p$ matrix function called the *coupling matrix* or *input matrix*. As a function of the state vector $\boldsymbol{x}$, $\boldsymbol{\mathcal{B}}$ maps from $\mathbb{R}^n$ to $\mathbb{R}^n$. A *single input* system has a scalar control signal $u(t)$, i.e., $p = 1$, and the coupling matrix $\boldsymbol{\mathcal{B}}(\boldsymbol{x})$ is a coupling vector written as $\boldsymbol{B}(\boldsymbol{x})$. The initial condition $\boldsymbol{x}_0$ prescribes the value of the state vector $\boldsymbol{x}$ at the initial time $t_0 \leq t$.

Regarding the state vector $\boldsymbol{x}$, the system Eq. (1.1) has two possible sources of nonlinearity. First, the nonlinearity $\boldsymbol{R}(\boldsymbol{x})$ typically is a nonlinear function of the state. This is the nonlinearity encountered in uncontrolled systems. Second, the

coupling matrix $\boldsymbol{\mathcal{B}}(\boldsymbol{x})$ may depend nonlinearly on the state $\boldsymbol{x}$. This nonlinearity is exclusive for control systems. Equation (1.1) is called an *affine control system* because the control signal $\boldsymbol{u}(t)$ enters only linearly. Throughout the thesis, it is assumed that the coupling matrix $\boldsymbol{\mathcal{B}}(\boldsymbol{x})$ has full rank for all values of $\boldsymbol{x}$. Because $p \leq n$, this condition is

$$\operatorname{rank}\left(\boldsymbol{\mathcal{B}}(\boldsymbol{x})\right) = p. \tag{1.4}$$

Assumption (1.4) ensures that the maximum number of p independent control signal acts on the system regardless of the value of the state vector $\boldsymbol{x}$.

1.2. Examples of affine control systems

Some examples of affine control systems are discussed. These examples are encountered repeatedly to illustrate the developed concepts.

Example 1.1: Mechanical control system in one spatial dimension

Newton's equation of motion for a single point mass in one spatial dimension x is (Goldstein et al., 2001),

$$\ddot{x}(t) = R\left(x(t), \dot{x}(t)\right) + B\left(x(t), \dot{x}(t)\right) u(t). \tag{1.5}$$

The point mass is moving in the external force field R which may depend on position x and velocity $\dot{x}$ of the particle. The control signal $u(t)$ couples to the point mass via the control force $B\left(x(t), \dot{x}(t)\right) u(t)$, with $u(t)$ being the control signal and B the coupling function. Introducing the velocity $y(t) = \dot{x}(t)$, Eq. (1.5) can be rearranged as an affine control system,

$$\dot{x}(t) = y(t), \tag{1.6}$$
$$\dot{y}(t) = R\left(x(t), y(t)\right) + B\left(x(t), y(t)\right) u(t). \tag{1.7}$$

In vector notation, Eqs. (1.6) and (1.7) become

$$\dot{\boldsymbol{x}}(t) = \boldsymbol{R}\left(\boldsymbol{x}(t)\right) + \boldsymbol{B}\left(\boldsymbol{x}(t)\right) u(t), \tag{1.8}$$

with

$$\boldsymbol{x}(t) = \left(x(t), y(t)\right)^T, \quad \boldsymbol{R}(\boldsymbol{x}) = \left(y, R(x,y)\right)^T, \quad \boldsymbol{B}(\boldsymbol{x}) = \left(0, B(x,y)\right)^T. \tag{1.9}$$

The condition of full rank for the coupling vector $\boldsymbol{B}$ is

$$\operatorname{rank}\left(\boldsymbol{B}(\boldsymbol{x})\right) = 1, \tag{1.10}$$

which in turn implies

$$B(x, y) \neq 0 \tag{1.11}$$

for all values of x and y. Note that the control force acts on the nonlinear equation (1.7) for the velocity $y(t)$, while the remaining equation (1.6) for the position $x(t)$ is linear.

Example 1.2: FitzHugh-Nagumo model

The FitzHugh-Nagumo (FHN) model (FitzHugh, 1961; Nagumo et al., 1962) is a simple nonlinear model describing a prototype excitable system (Izhikevich, 2010). It arose as a simplified version of the Hodgkin-Huxley model which describes action-potential dynamics in neurons (Hodgkin and Huxley, 1952; Keener and Sneyd, 2008a) and contains the Van der Pol oscillator as a special case (Van der Pol, 1926). The model consists of two variables called the inhibitor x and activator y,

$$\begin{pmatrix} \dot{x}(t) \\ \dot{y}(t) \end{pmatrix} = \begin{pmatrix} a_0 + a_1 x(t) + a_2 y(t) \\ R(x(t), y(t)) \end{pmatrix} + \boldsymbol{\mathcal{B}}(\boldsymbol{x}(t))\, \boldsymbol{u}(t). \tag{1.12}$$

The function $R(x, y)$ is given by

$$R(x, y) = R(y) - x, \tag{1.13}$$

with $R(y)$ being a cubic polynomial of the form

$$R(y) = y - \frac{1}{3}y^3. \tag{1.14}$$

The nonlinearity is linear in the inhibitor x but nonlinear in the activator y. Unless otherwise announced, a set of standard parameter values

$$a_0 = 0.056, \qquad a_1 = -0.064, \qquad a_2 = 0.08 \tag{1.15}$$

is used for numerical simulations. Because this model is not a mechanical system, it is not predefined in which way a control acts on the system. Several simple choices with a constant coupling matrix $\boldsymbol{\mathcal{B}}(\boldsymbol{x}) = \boldsymbol{\mathcal{B}}$ are possible. A control acting on the activator equation leads to a coupling vector $\boldsymbol{B} = \begin{pmatrix} 0, & 1 \end{pmatrix}^T$, while a control acting on the inhibitor equation gives $\boldsymbol{B} = \begin{pmatrix} 1, & 0 \end{pmatrix}^T$. The former is called the *activator-controlled FitzHugh-Nagumo model*, while the latter is named *inhibitor-controlled FitzHugh-Nagumo model*. The simplest case occurs if the number of independent control signals equals the number of state

components, $p = n$, and the coupling matrix $\boldsymbol{\mathcal{B}}$ attains the form

$$\boldsymbol{\mathcal{B}} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}. \tag{1.16}$$

Note that Eq. (1.12) reduces to a mechanical control system in one spatial dimension with external force $R(x, y)$ for the parameter values $a_0 = a_1 = 0$, and $a_2 = 1$, and a coupling vector $\boldsymbol{B}(\boldsymbol{x}) = \begin{pmatrix} 0, & B(x, y) \end{pmatrix}^T$.

Strictly speaking, only a model with $R(x, y)$ given by a cubic polynomial in y and linear in x is called the FitzHugh-Nagumo. The approach to control developed here identifies Eq. (1.12) with arbitrary nonlinearity $R(x, y)$ and coupling vector $\boldsymbol{B} = \begin{pmatrix} 0, & 1 \end{pmatrix}^T$ as a particularly simple form of two-dimensional controlled dynamical systems. In absence of a better name, this model is occasionally called the activator-controlled FHN model as well.

Example 1.3: SIR model

The SIR-model is a nonlinear dynamical system to describe the transmission of a disease among a population (Bailey, 1975; Murray, 2007, 2011). The original model was created by Kermack and McKendrick in 1927 (Kermack and McKendrick, 1927) and consists of three components

$$\dot{S}(t) = -\beta \frac{S(t) I(t)}{N}, \tag{1.17}$$

$$\dot{I}(t) = \beta \frac{S(t) I(t)}{N} - \gamma I(t), \tag{1.18}$$

$$\dot{R}(t) = \gamma I(t). \tag{1.19}$$

The variable S denotes the number of susceptible individuals. If susceptible individuals S come in contact with infected individuals I, they become infected with a transmission rate $\beta = 0.36$. The average period of infectiousness is set to $1/\gamma = 5$ days, after which infected individuals either recover or die. Both possibilities are collected in the variable R. Recovered or dead individuals R are immune and do not become susceptible again. The total population number $N = S(t) + I(t) + R(t)$ is constant in time because

$$\dot{S}(t) + \dot{I}(t) + \dot{R}(t) = 0. \tag{1.20}$$

No exact analytical solution to Eqs. (1.17)-(1.19) is known. Figure 1.1 shows a typical time evolution of an epidemic obtained by numerical simulations.

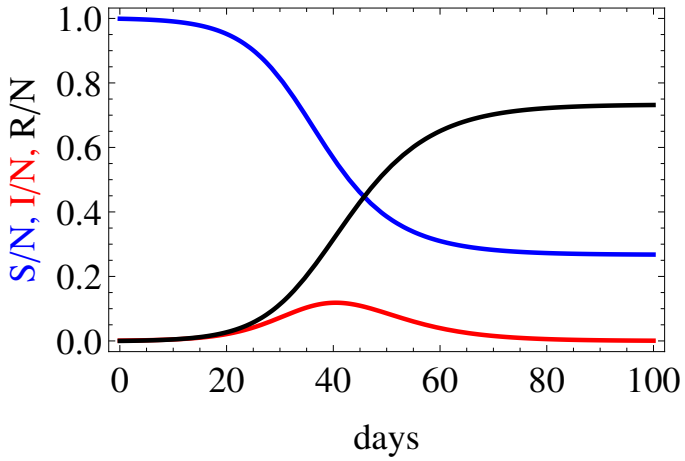

Figure 1.1.: Time evolution of an epidemic according to the SIR model. Time is measured in days. Initially, almost all individuals are susceptible (blue) and only very few are infected (red). The number of infected individuals reaches a maximum and subsequently decays to zero, with only susceptible and recovered individuals (black) remaining.

The reproductive number R_0 defined as

$$R_0 = \frac{\beta}{\gamma} \tag{1.21}$$

is the average number of susceptible individuals an infectious individual is infecting. To prevent further spreading of the epidemics, this number must be $R_0 < 1$.

There are two parameters β and γ in the system which can be affected by control measures. Culling of infected domestic animals to increase the rate γ is a common procedure but it is out of question for humans. A common measure during an epidemic among humans is to separate the infected persons from the susceptible persons. In the framework of the SIR model, such measures decrease the transmission rate $\beta(t)$. Thus, the transmission rate becomes time dependent and is of the form

$$\beta(t) = \beta + u(t). \tag{1.22}$$

Here, β is the constant transmission rate of the uncontrolled system and $u(t)$ is the control signal. The controlled SIR model investigated in this thesis is

$$\dot{\boldsymbol{x}}(t) = \boldsymbol{R}(\boldsymbol{x}(t)) + \boldsymbol{B}(\boldsymbol{x}(t))\, u(t), \tag{1.23}$$

with

$$\boldsymbol{x}(t) = (S(t), I(t), R(t))^T, \tag{1.24}$$

$$\boldsymbol{R}(\boldsymbol{x}(t)) = \left(-\beta\frac{S(t)\,I(t)}{N}, \beta\frac{S(t)\,I(t)}{N} - \gamma I(t), \gamma I(t)\right)^T, \tag{1.25}$$

$$\boldsymbol{B}(\boldsymbol{x}(t)) = \left(-\frac{S(t)\,I(t)}{N}, \frac{S(t)\,I(t)}{N}, 0\right)^T. \tag{1.26}$$

1.3. Optimal trajectory tracking

An important control objective is the guidance of state trajectories of dynamical systems along a desired reference trajectory. The reference trajectory is called the *desired trajectory* and denoted as $\boldsymbol{x}_d(t) \in \mathbb{R}^n$. It has the same number n of components as the system's state $\boldsymbol{x}(t)$ and is defined for a time interval $t_0 \leq t \leq t_1$. The closer the controlled state trajectory $\boldsymbol{x}(t)$ follows the reference trajectory $\boldsymbol{x}_d(t)$, the better the control target is achieved. In the ideal case, the desired trajectory $\boldsymbol{x}_d(t)$ is exactly equal to the actual controlled state trajectory $\boldsymbol{x}(t)$. A convenient measure for the distance between a desired trajectory $\boldsymbol{x}_d(t)$ and the actual trajectory $\boldsymbol{x}(t)$ of the controlled system is the squared difference integrated over the time interval,

$$\mathcal{J}[\boldsymbol{x}(t)] = \frac{1}{2}\int_{t_0}^{t_1} dt\,(\boldsymbol{x}(t) - \boldsymbol{x}_d(t))^2. \tag{1.27}$$

The quantity $\mathcal{J}[\boldsymbol{x}(t)]$ defined in Eq. (1.27) is a functional of the state vector $\boldsymbol{x}(t)$ over the time interval $t \in [t_0, t_1]$. It defines a distance between trajectories, i.e., a distance in function space. Optimal trajectory tracking aims to find the control signal $\boldsymbol{u}(t)$ such that $\mathcal{J}[\boldsymbol{x}(t)]$ is minimal. If this control solution exists and is unique, no "better" control signal exists. Any other control would result in a controlled state trajectory $\boldsymbol{x}(t)$ with a larger distance to the desired trajectory $\boldsymbol{x}_d(t)$. However, as will be discussed later on, the functional as defined in Eq. (1.27) might lead to an ill-defined control signal $\boldsymbol{u}(t)$ and state trajectory $\boldsymbol{x}(t)$. A possible remedy is to introduce a regularization term

$$\mathcal{J}[\boldsymbol{x}(t), \boldsymbol{u}(t)] = \frac{1}{2}\int_{t_0}^{t_1} dt\,(\boldsymbol{x}(t) - \boldsymbol{x}_d(t))^2 + \frac{\epsilon^2}{2}\int_{t_0}^{t_1} dt\,|\boldsymbol{u}(t)|^2. \tag{1.28}$$

The coefficient ϵ is a regularization parameter. The regularization term penalizes large controls and guarantees a well-defined solution to the optimization problem.

Mathematically speaking, the problem of finding the optimal control signal $\boldsymbol{u}(t)$ by minimizing Eq. (1.28) is a constrained minimization problem, with $\boldsymbol{x}(t)$ constrained

to be the solution to the controlled dynamical system (1.1) with initial condition (1.2). The standard approach to solving constrained optimization problems is to introduce Lagrange multipliers $\boldsymbol{\lambda}(t)$. The *co-state* $\boldsymbol{\lambda}(t)$ has the same number of components as the state $\boldsymbol{x}(t)$. Its time evolution is governed by the *adjoint equation*. In contrast to solving an uncontrolled problem, which only involves finding a solution for the state $\boldsymbol{x}(t)$ in state space with dimension n, solving an optimal control problem involves finding a solution to the coupled state and adjoint equations in the *extended state space* with dimension $2n$. This renders optimal control problems much more difficult than uncontrolled problems. Numerical solutions of optimal control problems suffer from inconvenient terminal conditions for the co-state and necessitate a computationally expensive iterative algorithm.

The next example shows optimal trajectory tracking in the FHN model. The solution is obtained numerically with the help of the open source package ACADO (Houska et al., 2013, 2011a,b).

Example 1.4: Optimal trajectory tracking in the FHN model

Optimal trajectory tracking is discussed for the activator-controlled FHN model of Example 1.2. The desired trajectory is chosen to be an ellipse,

$$x_d(t) = A_x \cos(2\pi t/T) - \frac{1}{2}, \qquad y_d(t) = A_y \sin(2\pi t/T) + \frac{1}{2}, \tag{1.29}$$

with $A_x = 1$, $A_y = 15$, and $T = 1$. The regularization parameter ϵ is set to the value $\epsilon = 10^{-3}$, such that the coefficient of the regularization term is $\sim 10^{-6}$. Within the time interval $0 = t_0 \leq t < t_1 = 1$, the controlled state shall follow the ellipse as closely as possible. The initial and terminal states lie exactly on the desired trajectory. Figure 1.2 shows a numerical solution of the optimal trajectory tracking problem. The controlled state trajectory $\boldsymbol{x}(t)$ (red dashed line) looks quite different from the desired trajectory $\boldsymbol{x}_d(t)$ (blue solid line). However, the control is optimal, and not other control would yield a controlled state trajectory closer to the desired trajectory as measured by the functional Eq. (1.28).

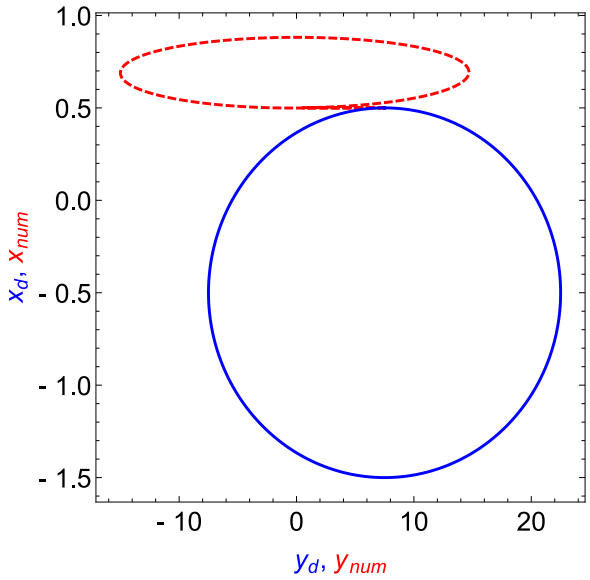

Figure 1.2.: Optimal trajectory tracking in the activator-controlled FHN model. The state space plot compares the desired trajectory (blue solid line) with the optimally controlled state trajectory (red dashed line). The agreement is not particularly impressive. However, the control is optimal, and no better control exists. Any other control yields a state trajectory $\boldsymbol{x}(t)$ with a larger distance to the desired trajectory $\boldsymbol{x}_d(t)$ as measured by the functional $\mathcal{J}$ defined in Eq. (1.28).

Comparing the individual state components with is desired counterparts in Fig. 1.2 somewhat clarifies the picture. The controlled activator (red dashed line in Fig. 1.2 right) is at least similar in shape to the desired activator (blue solid line) but seems to be shifted by a constant. A very steep initial transition leads from the initial condition onto the shifted trajectory, while a similarly steep transition occurs at the terminal time. The controlled inhibitor (red dashed line in Fig. 1.2 left) does not show any similarity to the desired inhibitor (blue solid line). In contrast to the activator component, it does not exhibit steep initial and terminal transitions.

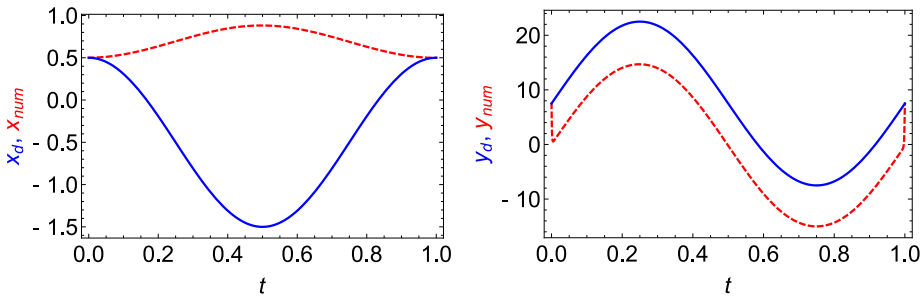

Figure 1.3.: Plotting the individual state components over time reveals a controlled activator component y similar in shape to y_d but shifted by a constant. Note the steep transition regions at the beginning and end of the time interval. However, the controlled inhibitor component x is very different from its desired counterpart x_d.

An important lesson is to be learned from Example 1.4. In general, it is impossible for the controlled state trajectory $\boldsymbol{x}(t)$ to follow exactly the desired trajectory $\boldsymbol{x}_d(t)$. The desired trajectory is that what you want, but it is usually not that what you get. If the numerical solution corresponds to the global minimum of the functional $\mathcal{J}$, Eq. (1.28), there is no other control which can enforce controlled state trajectory $\boldsymbol{x}(t)$ closer to $\boldsymbol{x}_d(t)$. Although the value of the functional $\mathcal{J}$ attains its minimally possible value, this value might still be very large, indicating a large distance between controlled and desired state trajectory. Naturally, the following question arises. Under which conditions is the controlled state trajectory $\boldsymbol{x}(t)$ identical to the desired trajectory $\boldsymbol{x}_d(t)$?

To answer that question, the concept of *exactly realizable trajectories* is proposed in Chapter 2. A desired trajectory is exactly realizable if it satisfies a condition called the *constraint equation*. An open loop control signal $\boldsymbol{u}(t)$ can be determined which enforces the state to follow the desired trajectory exactly, $\boldsymbol{x}(t) = \boldsymbol{x}_d(t)$. This approach does not only yield an explicit expression for the control signal in terms of the desired trajectory, but also identifies a particularly simple class of nonlinear affine control systems. Systems in this class share many properties with linear control systems and satisfy the so-called *linearizing assumption*. Chapter 3 relates exactly realizable trajectories to optimal control. In particular, an exactly realizable trajectories, together with the corresponding control signal, is the solution to an unregularized optimal control problem. Based on that insight, the regularization parameter ϵ is used as the small parameter for a singular perturbation expansion in Chapter 4. This results in a reinterpretation of affine optimal trajectory tracking problems with small regularization term as a system of singularly perturbed differential equations. Combining this approach with the linearizing assumption, approximate solutions for optimal trajectory tracking in terms of mostly linear equations can be derived. The analytical solutions are valid for arbitrary desired trajectories. This approach applies, among other systems, to the mechanical control systems from Example 1.2 and the activator-controlled FHN model from Example 1.4. Note that

the small parameter ϵ originates from the formulation of the control problem Eq. (1.28). Assuming this parameter to be small does *not* involve any approximations of the system dynamics. The system dynamics is exactly taken into account by the perturbative approach. While the analytical results are obtained for open loop control, they are modified in Section 4.3 to yield solutions for optimal feedback control. Chapter 5 extends the notion of exactly realizable trajectories to reaction-diffusion systems.

As a teaser and to demonstrate the accuracy of the analytical approximation, we compare the numerical solution for the optimal trajectory tracking in the FHN model from Example 1.4 with the analytical approximation in Example 1.5. The exact analytical expression and its derivation is quite involved. All details can be found in Chapter 4.

Example 1.5: Analytical approximation for the optimally controlled FHN model

Optimal trajectory tracking in the activator-controlled FHN model (see Example 1.2) can be approximately solved with the analytical techniques developed in this thesis. The regularization parameter ϵ in Eq. (1.28) is used as the small parameter for a singular perturbation expansion. For the same desired trajectories as in Example 1.4, and the same value of the regularization parameter $\epsilon = 10^{-3}$, Fig. 1.4 compares the analytical approximation for the optimally controlled state trajectory $\boldsymbol{x}(t)$ with the corresponding numerical solution. For such a small value of the regularization parameter $\epsilon = 10^{-3}$, the agreement is almost perfect.

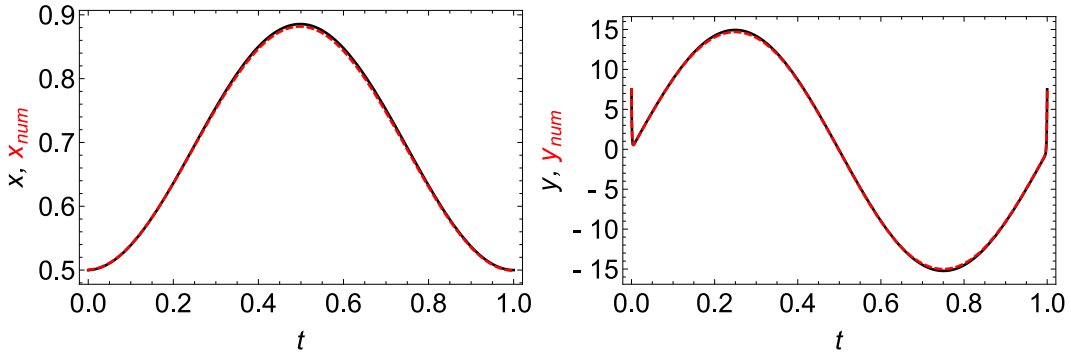

Figure 1.4.: Comparison of analytical approximation (black solid line) and numerically obtained (red dashed line) optimally controlled state trajectory of Example 1.4 for the activator y (right) and inhibitor x (left) over time.

The analytical result of Chapter 4 reveals a surprising result. The analytical approximation for the controlled state trajectory $\boldsymbol{x}(t)$ of Example 1.4 does not depend on the nonlinearity $R(x, y)$ (see Example 1.2 for the model equations)! More precisely,

changing the parameter values of the nonlinearity $R(x, y)$, or changing $R(x, y)$ altogether, has no effect on the controlled state trajectory. Although the system dynamics is governed by nonlinear differential equations, the optimally controlled system can be approximated by solving only linear equations. However, the analytical solution for the control signal depends strongly on the nonlinearity R. This prediction is verified with an additional numerical computation for a FHN like model with vanishing nonlinearity $R \equiv 0$ in Example 1.6.

Example 1.6: Optimal trajectory tracking for a linear system

The affine control system

$$\begin{pmatrix} \dot{x}(t) \\ \dot{y}(t) \end{pmatrix} = \begin{pmatrix} a_0 + a_1 x(t) + a_2 y(t) \\ 0 \end{pmatrix} + \begin{pmatrix} 0 \\ 1 \end{pmatrix} u(t), \tag{1.30}$$

has the same form as the FHN model of Example 1.2 except for the vanishing nonlinearity,

$$R(x, y) = 0. \tag{1.31}$$

The parameter values for a_0, a_1, and a_2 are the same as in Example 1.2. The analytical solution predicts that, in the limit of small regularization parameter $\epsilon \to 0$, the optimally controlled state trajectory is independent of the actual form of the nonlinearity R. Therefore, choosing the same desired trajectories Eq. (1.29) for the optimal control of Eq. (1.30) should yield the same optimally controlled state trajectories $\boldsymbol{x}(t)$ as in Example 1.4. Indeed, no discernible difference is visible in the numerical solutions for both problems, as Fig. 1.5 shows. However, the corresponding control signals depend on the nonlinearity, as is shown in Fig. 1.6.

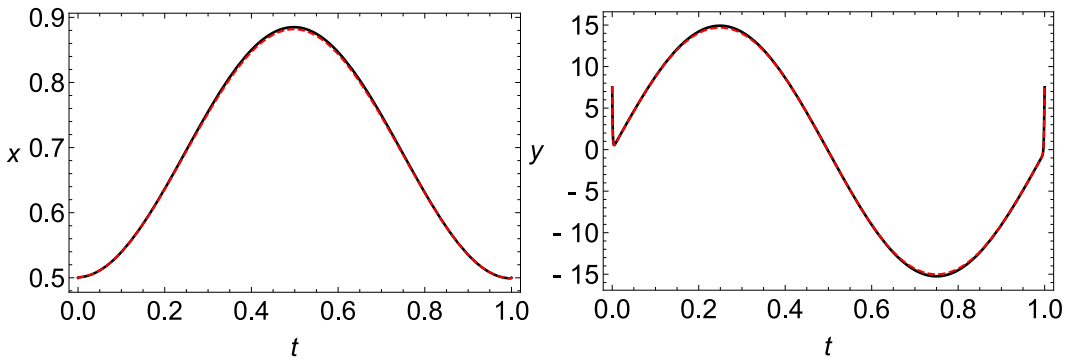

Figure 1.5.: The optimally controlled state trajectory $\boldsymbol{x} = (x, y)^T$ does not depend on the nonlinearity R in the limit of vanishing regularization parameter, $\epsilon \to 0$. The black solid line is the numerical result for a vanishing nonlinearity $R(x, y) = 0$, while the red dashed line shows the result for the FHN nonlinearity $R(x, y) = y - \frac{1}{3}y^3 - x$.

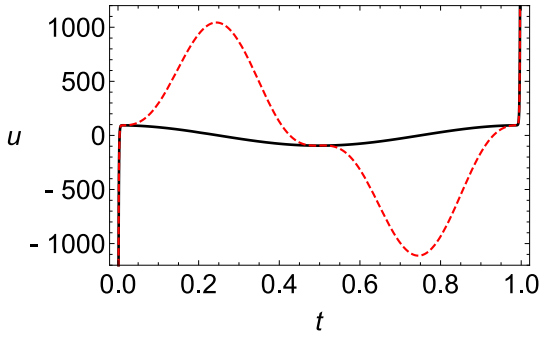

Figure 1.6.: The control signal u depends on the nonlinearity R. Black solid line is the numerical result for vanishing nonlinearity $R(x, y) = 0$, while the red dashed line shows the result for the standard FHN nonlinearity $R(x, y) = y - \frac{1}{3}y^3 - x$. Note that the control exhibits very large values and steep slopes at the beginning and end of the time interval in both cases.

The last example reveals that, under certain conditions, the nonlinearity $R(x, y)$ plays only a minor role for the controlled state trajectory. In fact, if the regularization parameter ϵ is zero, the approximate solution for the state trajectory is exact and governed solely by linear equations. The analytical treatment uncovers an underlying linear structure of certain nonlinear optimal trajectory tracking problems. However, this exact solution to nonlinear optimal control systems does not come without a price. For $\epsilon = 0$, the optimally controlled state trajectory cannot be expressed in terms of continuous functions, but involves jumps in at least one component. These jumps are located at the beginning and the end of the time interval. Even worse, the corresponding control signal does diverge at the same points. This behavior can already be anticipated from Fig. 1.5 right: the y-component of the state exhibits a steep transition region close to the beginning and the end of the time interval for a small but finite value of $\epsilon = 10^{-3}$. These transition regions degenerate

to jumps in the limit $\epsilon \to 0$. In the language of singular perturbation theory, such transition regions are known as boundary layers. Similarly, Fig. 1.6 shows that the corresponding control signal $u(t)$ exhibits very large amplitudes located equally at the beginning and the end of the time interval.

An underlying linear structure of nonlinear control systems might sound surprising. However, instances of exact linearizations are well known from mathematical control theory. A prominent example is feedback linearization, see e.g. (Khalil, 2001; Slotine and Li, 1991; Isidori, 1995). As a very simple example system, consider the activator-controlled FHN model,

$$\begin{pmatrix} \dot{x}(t) \\ \dot{y}(t) \end{pmatrix} = \begin{pmatrix} a_0 + a_1 x(t) + a_2 y(t) \\ R(x(t), y(t)) \end{pmatrix} + \begin{pmatrix} 0 \\ 1 \end{pmatrix} u(t), \tag{1.32}$$

with nonlinearity

$$R(x, y) = y - \frac{1}{3}y^3 - x. \tag{1.33}$$

As the name implies, feedback linearization assumes a feedback control from the very beginning, i.e., u may depend on state $\boldsymbol{x}$ as

$$u(t) = u(\boldsymbol{x}(t), t). \tag{1.34}$$

With the help of a very simple transform of the control signal, it is possible to obtain a controlled system linear in state and control. Introducing a new control signal $v(t)$ as

$$u(x(t), y(t), t) = -R(x(t), y(t)) + v(t), \tag{1.35}$$

the controlled system transforms to

$$\dot{x}(t) = y(t), \qquad \dot{y}(t) = v(t). \tag{1.36}$$

While the original control signal $u(t)$ depends on the nonlinearity R, the controlled state trajectory $\boldsymbol{x}(t)$, obtained as the solution to Eq. (1.36), does not depend on R. This is in fact similar to our approach.

In contrast to approximate linearizations performed to study the linear stability of solutions, feedback linearization is an exact transformation of a nonlinear to a linear dynamical system. Exact linearizations of uncontrolled dynamical systems exist as well. For example, a nonlinear transformation of the state converts Riccati equations to linear differential equations (Zaitsev and Polyanin, 2002). However, because the class of exactly linearizable uncontrolled systems is small, this method is rarely applied in practice. In general, feedback linearization applies a combined transformation of state and control to obtain a linear system. The class of feedback linearizable nonlinear control systems is huge, and the simple model Eq. (1.32) is only the trivial case requiring no state transformation (Khalil, 2001).

A disadvantage of feedback linearization is that it assumes a feedback control from the very beginning and does not easily apply to open loop control methods. Furthermore, feedback control might lead to results which are not optimal. In principle, a feedback-controlled nonlinear system can be much simpler than its corresponding uncontrolled counterpart. This is in contrast to optimally controlled systems, which are usually much more difficult than its uncontrolled counterpart due to the coupling of the state and co-state equations. As demonstrated by the example above, in this thesis we develop analytical techniques which reveal an underlying linear structure for a certain class of nonlinear optimal control systems. Similar to feedback linearization, the nonlinearity is absorbed by the control signal, and the time evolution of the controlled state trajectory is entirely determined by linear equations. These techniques apply only to a limited class of nonlinear control systems, and are not as powerful as feedback linearization. Nevertheless, this class includes some simple but important models motivated by physics and nonlinear dynamics, as the activator-controlled FHN model and mechanical control systems in one spatial dimension.

The chapter is concluded with a more philosophical remark. Trajectory tracking is actually ill-defined because it is a circular task. To achieve the aim of trajectory tracking, an appropriate control signal must be applied to the dynamical system. In a universe which consists exclusively of dynamical systems, this control signal must be the output of a dynamical system. The only way to obtain an output which behaves in exactly the way necessary for trajectory tracking is to control the dynamical system which generates the output. To perform the task of trajectory tracking, it is necessary to have a second system for which the task of trajectory tracking is already performed with sufficient accuracy. Trajectory tracking is a circular task.

2. Exactly realizable trajectories

This chapter introduces the notion of exactly realizable trajectories. The necessary formalism is established in Section 2.1. After the definition of exactly realizable trajectories in Section 2.2, the linearizing assumption is introduced in Section 2.3. This assumption defines a class of nonlinear control systems which, to a large extent, behave like linear control systems. Combining the notion of an exactly realizable trajectory with the linearizing assumption allows one to extend some well known results about the controllability of linear systems to nonlinear control systems in Sections 2.4 and 2.5. Output Realizability is discussed in Section 2.6, and Section 2.7 concludes with a discussion and outlook.

2.1. Formalism

This section introduces the formalism which is repeatedly used throughout the thesis. The main elements are two complementary projection matrices $\boldsymbol{\mathcal{P}}$ and $\boldsymbol{\mathcal{Q}}$. Projectors are a useful ingredient for a number of physical theories. Take, for example, quantum mechanics, which describes measurements as projections of the state (an element from a Hilbert space) onto a ray or unions of rays of the Hilbert space (Fick, 1988; Cohen-Tannoudji et al., 2010). Also in non-equilibrium statistical mechanics, projectors have found widespread application to separate a subsystem of interest from its bath (Balescu, 1975; Grabert, 1982).

To the best of our knowledge, projectors have not been utilized in the context of control systems. Section 2.1.1 defines the projectors and Section 2.1.2 separates the controlled state equation in two equations. The first equation involves the control signal, while the second equation is independent of the control signal. This formalism provides a useful approach for analyzing general affine control systems. Appendix A.2 demonstrates how projectors arise in the context of overdetermined and underdetermined systems of linear equations.

2.1.1. The projectors $\mathcal{P}$ and $\mathcal{Q}$

Consider the affine control system with state dependent coupling matrix $\boldsymbol{\mathcal{B}}(\boldsymbol{x}(t))$

$$\dot{\boldsymbol{x}}(t) = \boldsymbol{R}(\boldsymbol{x}(t)) + \boldsymbol{\mathcal{B}}(\boldsymbol{x}(t))\,\boldsymbol{u}(t)\,. \tag{2.1}$$

Define two complementary projectors $\boldsymbol{\mathcal{P}}$ and $\boldsymbol{\mathcal{Q}}$ in terms of the coupling matrix $\boldsymbol{\mathcal{B}}(\boldsymbol{x}(t))$ as

$$\boldsymbol{\mathcal{P}}(\boldsymbol{x}) = \boldsymbol{\mathcal{B}}(\boldsymbol{x}) \left(\boldsymbol{\mathcal{B}}^T(\boldsymbol{x}) \boldsymbol{\mathcal{B}}(\boldsymbol{x})\right)^{-1} \boldsymbol{\mathcal{B}}^T(\boldsymbol{x}), \tag{2.2}$$

$$\boldsymbol{\mathcal{Q}}(\boldsymbol{x}) = \mathbf{1} - \boldsymbol{\mathcal{P}}(\boldsymbol{x}). \tag{2.3}$$

$\boldsymbol{\mathcal{P}}$ and $\boldsymbol{\mathcal{Q}}$ are $n \times n$ matrices which, in general, do depend on the state $\boldsymbol{x}$. Note that the $p \times p$ matrix $\boldsymbol{\mathcal{B}}^T(\boldsymbol{x}) \boldsymbol{\mathcal{B}}(\boldsymbol{x})$ has full rank p because of assumption Eq. (1.4) that $\boldsymbol{\mathcal{B}}(\boldsymbol{x})$ has full rank. Therefore, $\boldsymbol{\mathcal{B}}^T(\boldsymbol{x}) \boldsymbol{\mathcal{B}}(\boldsymbol{x})$ is a quadratic and non-singular matrix and its inverse exists. The projectors $\boldsymbol{\mathcal{P}}$ and $\boldsymbol{\mathcal{Q}}$ are also known as *Moore-Penrose projectors*. The rank of $\boldsymbol{\mathcal{P}}(\boldsymbol{x})$ and $\boldsymbol{\mathcal{Q}}(\boldsymbol{x})$ is

$$\operatorname{rank}(\boldsymbol{\mathcal{P}}(\boldsymbol{x})) = p, \qquad \operatorname{rank}(\boldsymbol{\mathcal{Q}}(\boldsymbol{x})) = n - p. \tag{2.4}$$

Multiplying the n-component state vector $\boldsymbol{x}$ by the $n \times n$ matrix $\boldsymbol{\mathcal{P}}(\boldsymbol{x})$ yields an n-component vector $\boldsymbol{z} = \boldsymbol{\mathcal{P}}(\boldsymbol{x}) \boldsymbol{x}$. However, because $\boldsymbol{\mathcal{P}}(\boldsymbol{x})$ has rank p, only p components of $\boldsymbol{z}$ are independent. Similar, only $n - p$ components of $\boldsymbol{y} = \boldsymbol{\mathcal{Q}}(\boldsymbol{x}) \boldsymbol{x}$ are independent.

From the definitions Eqs. (2.2) and (2.3) follow the projector properties idempotence

$$\boldsymbol{\mathcal{Q}}(\boldsymbol{x}) \boldsymbol{\mathcal{Q}}(\boldsymbol{x}) = \boldsymbol{\mathcal{Q}}(\boldsymbol{x}), \qquad \boldsymbol{\mathcal{P}}(\boldsymbol{x}) \boldsymbol{\mathcal{P}}(\boldsymbol{x}) = \boldsymbol{\mathcal{P}}(\boldsymbol{x}), \tag{2.5}$$

and complementarity

$$\boldsymbol{\mathcal{Q}}(\boldsymbol{x}) \boldsymbol{\mathcal{P}}(\boldsymbol{x}) = \boldsymbol{\mathcal{P}}(\boldsymbol{x}) \boldsymbol{\mathcal{Q}}(\boldsymbol{x}) = \mathbf{0}. \tag{2.6}$$

The projectors are symmetric,

$$\boldsymbol{\mathcal{P}}^T(\boldsymbol{x}) = \boldsymbol{\mathcal{P}}(\boldsymbol{x}), \qquad \boldsymbol{\mathcal{Q}}^T(\boldsymbol{x}) = \boldsymbol{\mathcal{Q}}(\boldsymbol{x}), \tag{2.7}$$

because the inverse of the symmetric matrix $\boldsymbol{\mathcal{B}}^T(\boldsymbol{x}) \boldsymbol{\mathcal{B}}(\boldsymbol{x})$ is symmetric. Furthermore, matrix multiplication from the right with the input matrix $\boldsymbol{\mathcal{B}}(\boldsymbol{x})$ yields the important relations

$$\boldsymbol{\mathcal{P}}(\boldsymbol{x}) \boldsymbol{\mathcal{B}}(\boldsymbol{x}) = \boldsymbol{\mathcal{B}}(\boldsymbol{x}), \qquad \boldsymbol{\mathcal{Q}}(\boldsymbol{x}) \boldsymbol{\mathcal{B}}(\boldsymbol{x}) = \mathbf{0}. \tag{2.8}$$

Similarly, matrix multiplication from the left with the transposed input matrix $\boldsymbol{\mathcal{B}}^T(\boldsymbol{x})$ yields

$$\boldsymbol{\mathcal{B}}^T(\boldsymbol{x}) \boldsymbol{\mathcal{P}}(\boldsymbol{x}) = \boldsymbol{\mathcal{B}}^T(\boldsymbol{x}), \qquad \boldsymbol{\mathcal{B}}^T(\boldsymbol{x}) \boldsymbol{\mathcal{Q}}(\boldsymbol{x}) = \mathbf{0}. \tag{2.9}$$

Some more properties of $\boldsymbol{\mathcal{P}}$ and $\boldsymbol{\mathcal{Q}}$ necessary for later chapters are compiled in Appendix A.3.

2.1.2. Separation of the state equation

The projectors defined in Eqs. (2.2) and (2.3) are used to split up the controlled state equation

$$\dot{\boldsymbol{x}}(t) = \boldsymbol{R}(\boldsymbol{x}(t)) + \boldsymbol{\mathcal{B}}(\boldsymbol{x}(t))\,\boldsymbol{u}(t). \tag{2.10}$$

Multiplying every term by $\mathbf{1} = \boldsymbol{\mathcal{P}}(\boldsymbol{x}(t)) + \boldsymbol{\mathcal{Q}}(\boldsymbol{x}(t))$, Eq. (2.10) can be written as

$$\begin{aligned}\frac{d}{dt}\left(\boldsymbol{\mathcal{P}}(\boldsymbol{x}(t))\,\boldsymbol{x}(t) + \boldsymbol{\mathcal{Q}}(\boldsymbol{x}(t))\,\boldsymbol{x}(t)\right) = &\left(\boldsymbol{\mathcal{P}}(\boldsymbol{x}(t)) + \boldsymbol{\mathcal{Q}}(\boldsymbol{x}(t))\right)\boldsymbol{R}(\boldsymbol{x}(t)) \\ &+ \left(\boldsymbol{\mathcal{P}}(\boldsymbol{x}(t)) + \boldsymbol{\mathcal{Q}}(\boldsymbol{x}(t))\right)\boldsymbol{\mathcal{B}}(\boldsymbol{x}(t))\,\boldsymbol{u}(t).\end{aligned} \tag{2.11}$$

Multiplying with $\boldsymbol{\mathcal{Q}}(\boldsymbol{x}(t))$ from the left and using Eq. (2.8) yields an equation independent of the control signal $\boldsymbol{u}$,

$$\boldsymbol{\mathcal{Q}}(\boldsymbol{x}(t))\left(\dot{\boldsymbol{x}}(t) - \boldsymbol{R}(\boldsymbol{x}(t))\right) = \mathbf{0}. \tag{2.12}$$

Equation (2.79) is called the *constraint equation*. Multiplying the controlled state equation (2.10) by $\boldsymbol{\mathcal{B}}^T(\boldsymbol{x}(t))$ from the left yields

$$\boldsymbol{\mathcal{B}}^T(\boldsymbol{x}(t))\,\dot{\boldsymbol{x}}(t) = \boldsymbol{\mathcal{B}}^T(\boldsymbol{x}(t))\,\boldsymbol{R}(\boldsymbol{x}(t)) + \boldsymbol{\mathcal{B}}^T(\boldsymbol{x}(t))\,\boldsymbol{\mathcal{B}}(\boldsymbol{x}(t))\,\boldsymbol{u}(t). \tag{2.13}$$

Multiplying with $\left(\boldsymbol{\mathcal{B}}^T(\boldsymbol{x}(t))\,\boldsymbol{\mathcal{B}}(\boldsymbol{x}(t))\right)^{-1}$, which exists as long as $\boldsymbol{\mathcal{B}}(\boldsymbol{x}(t))$ has full rank, from the left results in an expression for the vector of control signals $\boldsymbol{u}(t)$ in terms of the controlled state trajectory $\boldsymbol{x}(t)$,

$$\boldsymbol{u}(t) = \boldsymbol{\mathcal{B}}^+(\boldsymbol{x}(t))\left(\dot{\boldsymbol{x}}(t) - \boldsymbol{R}(\boldsymbol{x}(t))\right). \tag{2.14}$$

The abbreviation

$$\boldsymbol{\mathcal{B}}^+(\boldsymbol{x}) = \left(\boldsymbol{\mathcal{B}}^T(\boldsymbol{x})\,\boldsymbol{\mathcal{B}}(\boldsymbol{x})\right)^{-1}\boldsymbol{\mathcal{B}}^T(\boldsymbol{x}) \tag{2.15}$$

is known as the *Moore-Penrose pseudo inverse* of the matrix $\boldsymbol{\mathcal{B}}(\boldsymbol{x})$ (Campbell and Meyer Jr., 1991). See also Appendix A.2 how to express a solution to an overdetermined system of linear equations in terms of the Moore-Penrose pseudo inverse. With the help of $\boldsymbol{\mathcal{B}}^+$, the projector $\boldsymbol{\mathcal{P}}$ can be expressed as

$$\boldsymbol{\mathcal{P}}(\boldsymbol{x}) = \boldsymbol{\mathcal{B}}(\boldsymbol{x})\left(\boldsymbol{\mathcal{B}}^T(\boldsymbol{x})\,\boldsymbol{\mathcal{B}}(\boldsymbol{x})\right)^{-1}\boldsymbol{\mathcal{B}}^T(\boldsymbol{x}) = \boldsymbol{\mathcal{B}}(\boldsymbol{x})\,\boldsymbol{\mathcal{B}}^+(\boldsymbol{x}). \tag{2.16}$$

Note that

$$\begin{aligned}\boldsymbol{\mathcal{B}}^+(\boldsymbol{x}(t))\,\dot{\boldsymbol{x}}(t) &= \left(\boldsymbol{\mathcal{B}}^T(\boldsymbol{x}(t))\,\boldsymbol{\mathcal{B}}(\boldsymbol{x}(t))\right)^{-1}\boldsymbol{\mathcal{B}}^T(\boldsymbol{x}(t))\,\dot{\boldsymbol{x}}(t) \\ &= \left(\boldsymbol{\mathcal{B}}^T(\boldsymbol{x}(t))\,\boldsymbol{\mathcal{B}}(\boldsymbol{x}(t))\right)^{-1}\boldsymbol{\mathcal{B}}^T(\boldsymbol{x}(t))\,\boldsymbol{\mathcal{P}}(\boldsymbol{x}(t))\,\dot{\boldsymbol{x}}(t),\end{aligned} \tag{2.17}$$

such that expression (2.14) for the control involves only the time derivative $\boldsymbol{\mathcal{P}}(\boldsymbol{x}(t))\,\dot{\boldsymbol{x}}(t)$ and does not depend on $\boldsymbol{\mathcal{Q}}(\boldsymbol{x}(t))\,\dot{\boldsymbol{x}}(t)$.

In conclusion, every affine controlled state equation (2.10) can be split in two equations. The equation (2.14) involving $\boldsymbol{\mathcal{P}}\dot{\boldsymbol{x}}$ determines the control signal $\boldsymbol{u}$ in terms of the controlled state trajectory $\boldsymbol{x}$ and its derivative. The constraint equation (2.12) involves only $\boldsymbol{\mathcal{Q}}\dot{\boldsymbol{x}}$ and does not depend on the control signal. These relations are valid for any kind of control, be it an open or a closed loop control. The proposed separation of the state equation plays a central role in this thesis.

To illustrate the approach, the separation of the state equation is discussed with the help of two simple examples.

Example 2.1: Mechanical control system in one spatial dimension

The controlled state equation for mechanical control systems is (see Example 1.1),

$$\begin{pmatrix} \dot{x}(t) \\ \dot{y}(t) \end{pmatrix} = \begin{pmatrix} y(t) \\ R(x(t), y(t)) \end{pmatrix} + \begin{pmatrix} 0 \\ B(x(t), y(t)) \end{pmatrix} u(t). \tag{2.18}$$

The 2×1 coupling matrix is a vector which depends on the state vector $\boldsymbol{x}(t) = \begin{pmatrix} x(t), & y(t) \end{pmatrix}$,

$$\boldsymbol{B}(\boldsymbol{x}) = \begin{pmatrix} 0 \\ B(x,y) \end{pmatrix}, \tag{2.19}$$

while its transpose is a row vector

$$\boldsymbol{B}^T(\boldsymbol{x}) = \begin{pmatrix} 0, & B(x,y) \end{pmatrix}. \tag{2.20}$$

The computation of the Moore-Penrose pseudo inverse $\boldsymbol{B}^+$ involves the inner product

$$\boldsymbol{B}^T(\boldsymbol{x})\,\boldsymbol{B}(\boldsymbol{x}) = \begin{pmatrix} 0, & B(x,y) \end{pmatrix} \begin{pmatrix} 0 \\ B(x,y) \end{pmatrix} = B(x,y)^2. \tag{2.21}$$

The pseudo inverse $\boldsymbol{B}^+$ of $\boldsymbol{B}$ is given by

$$\boldsymbol{B}^+(\boldsymbol{x}) = \left(\boldsymbol{B}^T(\boldsymbol{x})\,\boldsymbol{B}(\boldsymbol{x})\right)^{-1} \boldsymbol{B}^T(\boldsymbol{x}) = B(x,y)^{-2} \begin{pmatrix} 0, & B(x,y) \end{pmatrix}, \tag{2.22}$$

while the projectors $\boldsymbol{\mathcal{P}}(\boldsymbol{x})$ and $\boldsymbol{\mathcal{Q}}(\boldsymbol{x})$ are given by

$$\begin{aligned}\boldsymbol{\mathcal{P}}(\boldsymbol{x}) &= \boldsymbol{\mathcal{P}} = \boldsymbol{B}(\boldsymbol{x})\left(\boldsymbol{B}^T(\boldsymbol{x})\,\boldsymbol{B}(\boldsymbol{x})\right)^{-1}\boldsymbol{B}^T(\boldsymbol{x}) \\ &= \begin{pmatrix} 0 \\ B(x,y) \end{pmatrix} B(x,y)^{-2} \begin{pmatrix} 0, & B(x,y) \end{pmatrix} = \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix}, \qquad (2.23)\end{aligned}$$

$$\boldsymbol{\mathcal{Q}}(\boldsymbol{x}) = \boldsymbol{\mathcal{Q}} = \mathbf{1} - \boldsymbol{\mathcal{P}}(\boldsymbol{x}) = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}. \qquad (2.24)$$

Although the coupling vector $\boldsymbol{B}(\boldsymbol{x})$ depends on the state $\boldsymbol{x}$, the projectors $\boldsymbol{\mathcal{P}}$ and $\boldsymbol{\mathcal{Q}}$ are actually independent of the state. With $\boldsymbol{\mathcal{P}}$ and $\boldsymbol{\mathcal{Q}}$, the state $\boldsymbol{x}$ can be split up in two parts,

$$\boldsymbol{\mathcal{P}}\boldsymbol{x}(t) = \begin{pmatrix} 0 \\ y(t) \end{pmatrix}, \qquad \boldsymbol{\mathcal{Q}}\boldsymbol{x}(t) = \begin{pmatrix} x(t) \\ 0 \end{pmatrix}. \qquad (2.25)$$

Both parts are vectors with two components, but have only one non-vanishing component. The control signal can be expressed in terms of the controlled state trajectory $\boldsymbol{x}(t)$ as

$$\begin{aligned}u(t) &= \boldsymbol{B}^{+}(\boldsymbol{x}(t))\left(\dot{\boldsymbol{x}}(t) - \boldsymbol{R}(\boldsymbol{x}(t))\right) \\ &= B(x(t), y(t))^{-2} \begin{pmatrix} 0, & B(x,y) \end{pmatrix} \left(\begin{pmatrix} \dot{x}(t) \\ \dot{y}(t) \end{pmatrix} - \begin{pmatrix} y(t) \\ R(x(t), y(t)) \end{pmatrix} \right) \\ &= \frac{1}{B(x(t), y(t))}\left(\dot{y}(t) - R(x(t), y(t))\right). \qquad (2.26)\end{aligned}$$

Note that the assumption of full rank for the coupling vector $\boldsymbol{B}$ implies that the function $B(x,y)$ does not vanish, i.e., $B(x,y) \neq 0$ for all values of x and y. Consequently, $u(t)$ is well defined for all times.

Example 2.2: Single input diagonal LTI system

Consider a diagonal 2×2 linear time-invariant (LTI) system for the state vector $\boldsymbol{x}(t) = \begin{pmatrix} x_1(t), & x_2(t) \end{pmatrix}^T$. Let both components be controlled by the same control signal $u(t)$,

$$\dot{x}_1(t) = \lambda_1 x_1(t) + u(t), \qquad \dot{x}_2(t) = \lambda_2 x_2(t) + u(t). \qquad (2.27)$$

The state matrix $\boldsymbol{\mathcal{A}}$ and input matrix $\boldsymbol{\mathcal{B}}$ are

$$\boldsymbol{\mathcal{A}} = \begin{pmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{pmatrix}, \qquad \boldsymbol{\mathcal{B}} = \begin{pmatrix} 1 \\ 1 \end{pmatrix}. \qquad (2.28)$$

The constant projectors $\boldsymbol{\mathcal{P}}$ and $\boldsymbol{\mathcal{Q}}$ can be computed as

$$\boldsymbol{\mathcal{P}} = \frac{1}{2}\begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}, \qquad \boldsymbol{\mathcal{Q}} = \frac{1}{2}\begin{pmatrix} 1 & -1 \\ -1 & 1 \end{pmatrix}. \tag{2.29}$$

The two projections of the state $\boldsymbol{x}(t)$ are

$$\boldsymbol{z}(t) = \begin{pmatrix} z_1(t) \\ z_2(t) \end{pmatrix} = \boldsymbol{\mathcal{P}}\boldsymbol{x}(t) = \frac{1}{2}\begin{pmatrix} x_1(t) + x_2(t) \\ x_1(t) + x_2(t) \end{pmatrix} \tag{2.30}$$

and

$$\boldsymbol{y}(t) = \begin{pmatrix} y_1(t) \\ y_2(t) \end{pmatrix} = \boldsymbol{\mathcal{Q}}\boldsymbol{x}(t) = \frac{1}{2}\begin{pmatrix} x_1(t) - x_2(t) \\ x_2(t) - x_1(t) \end{pmatrix}. \tag{2.31}$$

While both components of $\boldsymbol{y}(t)$ are non-zero, they are not linearly independent. The component $y_2(t)$ is redundant and is simply given by

$$y_2(t) = -y_1(t). \tag{2.32}$$

Similarly, the component $z_2(t)$ of vector $\boldsymbol{z}(t)$ is redundant because of

$$z_2(t) = z_1(t). \tag{2.33}$$

Example 2.2 shows that the projectors $\boldsymbol{\mathcal{P}}$ and $\boldsymbol{\mathcal{Q}}$ do not necessarily project onto single components of the state vector. If the projectors $\boldsymbol{\mathcal{P}}(\boldsymbol{x}) = \boldsymbol{\mathcal{P}}$ and $\boldsymbol{\mathcal{Q}}(\boldsymbol{x}) = \boldsymbol{\mathcal{Q}}$ are independent of the state $\boldsymbol{x}$, the parts $\boldsymbol{z} = \boldsymbol{\mathcal{P}}\boldsymbol{x}$ and $\boldsymbol{y} = \boldsymbol{\mathcal{Q}}\boldsymbol{x}$ are linear combinations of the original state components $\boldsymbol{x}$. If $\boldsymbol{\mathcal{P}} = \boldsymbol{\mathcal{P}}(\boldsymbol{x})$ and therefore also $\boldsymbol{\mathcal{Q}}(\boldsymbol{x}) = \mathbf{1} - \boldsymbol{\mathcal{P}}(\boldsymbol{x})$ depend on the state $\boldsymbol{x}$ itself, both parts $\boldsymbol{y}$ and $\boldsymbol{z}$ are nonlinear functions of the state $\boldsymbol{x}$. Only if the projectors are diagonal, constant, and appropriately ordered, $\boldsymbol{\mathcal{P}}(\boldsymbol{x}) = \boldsymbol{\mathcal{P}}_D$ and $\boldsymbol{\mathcal{Q}}(\boldsymbol{x}) = \boldsymbol{\mathcal{Q}}_D$, then the two parts $\boldsymbol{y}$ and $\boldsymbol{z}$ attain the particularly simple form

$$\boldsymbol{y} = \boldsymbol{\mathcal{Q}}_D\boldsymbol{x} = \begin{pmatrix} 0, & \ldots, & 0, & x_{p+1}, & \ldots, & x_n \end{pmatrix}^T, \tag{2.34}$$

$$\boldsymbol{z} = \boldsymbol{\mathcal{P}}_D\boldsymbol{x} = \begin{pmatrix} x_1, & \ldots, & x_p, & 0, & \ldots, & 0 \end{pmatrix}^T. \tag{2.35}$$

Only this form allows a clear interpretation which component of $\boldsymbol{x}$ belongs to which part. However, in any case, $\boldsymbol{y} = \boldsymbol{\mathcal{Q}}(\boldsymbol{x})\,\boldsymbol{x}$ has exactly $n-p$ independent components because $\boldsymbol{\mathcal{Q}}(\boldsymbol{x})$ has rank $n-p$, while $\boldsymbol{z} = \boldsymbol{\mathcal{P}}(\boldsymbol{x})\,\boldsymbol{x}$ has p independent components because $\boldsymbol{\mathcal{P}}(\boldsymbol{x})$ has rank p.

Projectors have only zeros and ones as possible eigenvalues. The diagonalization of an $n \times n$ projector $\boldsymbol{\mathcal{P}}(\boldsymbol{x})$ with $\operatorname{rank}(\boldsymbol{\mathcal{P}}(\boldsymbol{x})) = p$ is always possible (Fischer, 2013) and results in a diagonal $n \times n$ matrix with p entries of value one and $n-p$ entries of value zero on the diagonal. The transformation of the projectors $\boldsymbol{\mathcal{P}}(\boldsymbol{x})$ and $\boldsymbol{\mathcal{Q}}(\boldsymbol{x})$ to their diagonal counterparts defines a transformation of the state $\boldsymbol{x}$. See

Appendix A.4 how to construct this transformation. The transformation is nonlinear if $\boldsymbol{\mathcal{P}}(\boldsymbol{x})$ and $\boldsymbol{\mathcal{Q}}(\boldsymbol{x})$ are state dependent. Expressed in terms of the transformed state, the projectors $\boldsymbol{\mathcal{P}}$ and $\boldsymbol{\mathcal{Q}}$ are constant, diagonal, and appropriately ordered. Consequently, they yield a state separation of the form Eqs. (2.34)-(2.35). Such a representation defines a normal form of an affine control system. For a specified affine control system, computations will usually be simpler after the system is transformed to its normal form. However, for computations with general affine control systems, it is dispensable to perform the transformation if $\boldsymbol{z}(\boldsymbol{x}) = \boldsymbol{\mathcal{P}}(\boldsymbol{x})\,\boldsymbol{x}$ and $\boldsymbol{y}(\boldsymbol{x}) = \boldsymbol{\mathcal{Q}}(\boldsymbol{x})\,\boldsymbol{x}$ are simply viewed as separate parts. This allows a coordinate-free treatment of affine control systems.

2.2. Exactly realizable trajectories

As demonstrated in Example 1.4, not every desired state trajectory $\boldsymbol{x}_d(t)$ can be realized by control. Here, we answer the question under which conditions a desired trajectory $\boldsymbol{x}_d(t)$ is exactly realizable.

Consider the controlled state equation

$$\dot{\boldsymbol{x}}(t) = \boldsymbol{R}(\boldsymbol{x}(t)) + \boldsymbol{\mathcal{B}}(\boldsymbol{x}(t))\,\boldsymbol{u}(t), \tag{2.36}$$

$$\boldsymbol{x}(t_0) = \boldsymbol{x}_0. \tag{2.37}$$

The notion of *exactly realizable trajectories* is introduced. A realizable trajectory is a desired trajectory $\boldsymbol{x}_d(t)$ which satisfies two conditions.

1. The desired trajectory $\boldsymbol{x}_d(t)$ satisfies the constraint equation

$$\boldsymbol{\mathcal{Q}}(\boldsymbol{x}_d(t))\,(\dot{\boldsymbol{x}}_d(t) - \boldsymbol{R}(\boldsymbol{x}_d(t))) = \boldsymbol{0}. \tag{2.38}$$

2. The initial value $\boldsymbol{x}_d(t_0)$ must equal the initial value $\boldsymbol{x}_0$ of the controlled state equation,

$$\boldsymbol{x}_d(t_0) = \boldsymbol{x}_0. \tag{2.39}$$

The control solution for an exactly realizable trajectory is given by

$$\boldsymbol{u}(t) = \boldsymbol{\mathcal{B}}^{+}(\boldsymbol{x}_d(t))\,(\dot{\boldsymbol{x}}_d(t) - \boldsymbol{R}(\boldsymbol{x}_d(t))), \tag{2.40}$$

with the Moore-Penrose pseudo inverse $p \times n$ matrix $\boldsymbol{\mathcal{B}}^{+}$ defined as

$$\boldsymbol{\mathcal{B}}^{+}(\boldsymbol{x}) = \left(\boldsymbol{\mathcal{B}}^{T}(\boldsymbol{x})\,\boldsymbol{\mathcal{B}}(\boldsymbol{x})\right)^{-1}\boldsymbol{\mathcal{B}}^{T}(\boldsymbol{x}). \tag{2.41}$$

The notion of an exactly realizable trajectory allows the proof of the following statement.

If $\boldsymbol{x}_d(t)$ is an exactly realizable trajectory, i.e., if $\boldsymbol{x}_d(t)$ satisfies both conditions 1 and 2, then the state trajectory $\boldsymbol{x}(t)$ follows the desired trajectory $\boldsymbol{x}_d(t)$ exactly,

$$\boldsymbol{x}(t) = \boldsymbol{x}_d(t). \tag{2.42}$$

Using the control solution Eq. (2.40) in the controlled state equation (2.36) yields the following equation for the controlled state

$$\dot{\boldsymbol{x}}(t) = \boldsymbol{R}(\boldsymbol{x}(t)) + \boldsymbol{\mathcal{B}}(\boldsymbol{x}(t))\,\boldsymbol{\mathcal{B}}^{+}(\boldsymbol{x}_d(t))\,(\dot{\boldsymbol{x}}_d(t) - \boldsymbol{R}(\boldsymbol{x}_d(t))). \tag{2.43}$$

Note that $\boldsymbol{\mathcal{B}}$ depends on the actual system state $\boldsymbol{x}(t)$ while $\boldsymbol{\mathcal{B}}^{+}$ depends on the desired trajectory $\boldsymbol{x}_d(t)$. The difference $\Delta\boldsymbol{x}(t)$ between the true state $\boldsymbol{x}(t)$ and the desired trajectory $\boldsymbol{x}_d(t)$ is defined as

$$\Delta\boldsymbol{x}(t) = \boldsymbol{x}(t) - \boldsymbol{x}_d(t). \tag{2.44}$$

Using the definition for $\Delta\boldsymbol{x}(t)$ and Eq. (2.43) results in an ordinary differential equation (ODE) for $\Delta\boldsymbol{x}(t)$,

$$\begin{aligned}\Delta\dot{\boldsymbol{x}}(t) &= \boldsymbol{R}(\Delta\boldsymbol{x}(t) + \boldsymbol{x}_d(t)) - \dot{\boldsymbol{x}}_d(t)\\ &\quad + \boldsymbol{\mathcal{B}}(\Delta\boldsymbol{x}(t) + \boldsymbol{x}_d(t))\,\boldsymbol{\mathcal{B}}^{+}(\boldsymbol{x}_d(t))\,(\dot{\boldsymbol{x}}_d(t) - \boldsymbol{R}(\boldsymbol{x}_d(t))), \qquad (2.45)\\ \Delta\boldsymbol{x}(t_0) &= \boldsymbol{x}(t_0) - \boldsymbol{x}_d(t_0). \qquad (2.46)\end{aligned}$$

Assuming $|\Delta\boldsymbol{x}(t)| \ll 1$ and expanding Eq. (2.45) in $\Delta\boldsymbol{x}(t)$ yields

$$\begin{aligned}\Delta\dot{\boldsymbol{x}}(t) &= \boldsymbol{R}(\boldsymbol{x}_d(t)) - \dot{\boldsymbol{x}}_d(t) + \boldsymbol{\mathcal{B}}(\boldsymbol{x}_d(t))\,\boldsymbol{\mathcal{B}}^{+}(\boldsymbol{x}_d(t))\,(\dot{\boldsymbol{x}}_d(t) - \boldsymbol{R}(\boldsymbol{x}_d(t)))\\ &\quad + \nabla\boldsymbol{R}(\boldsymbol{x}_d(t))\,\Delta\boldsymbol{x}(t) + (\nabla\boldsymbol{\mathcal{B}}(\boldsymbol{x}_d(t))\,\Delta\boldsymbol{x}(t))\,\boldsymbol{\mathcal{B}}^{+}(\boldsymbol{x}_d(t))\,(\dot{\boldsymbol{x}}_d(t) - \boldsymbol{R}(\boldsymbol{x}_d(t)))\\ &\quad + \mathcal{O}\left(\Delta\boldsymbol{x}(t)^2\right). \qquad (2.47)\end{aligned}$$

Note that assuming $|\Delta\boldsymbol{x}(t)| \ll 1$ and subsequently expanding in $\Delta\boldsymbol{x}(t)$ does not result in a loss of generality of the final outcome. The expression $\nabla\boldsymbol{R}(\boldsymbol{x})$ denotes the Jacobian matrix of the nonlinearity $\boldsymbol{R}(\boldsymbol{x})$ with components

$$(\nabla\boldsymbol{R}(\boldsymbol{x}))_{ij} = \frac{\partial}{\partial x_j} R_i(\boldsymbol{x}),\ i,j \in \{1,\ldots,n\}. \tag{2.48}$$

The Jacobian $\nabla\boldsymbol{\mathcal{B}}(\boldsymbol{x})$ of $\boldsymbol{\mathcal{B}}$ is a third order tensor with components

$$(\nabla\boldsymbol{\mathcal{B}}(\boldsymbol{x}))_{ijk} = \frac{\partial}{\partial x_k}\mathcal{B}_{ij}(\boldsymbol{x}),\ i,k \in \{1,\ldots,n\},\ j \in \{1,\ldots p\}. \tag{2.49}$$

In the first line of Eq. (2.47), one can recognize the projector $\boldsymbol{\mathcal{B}}(\boldsymbol{x}_d(t))\,\boldsymbol{\mathcal{B}}^{+}(\boldsymbol{x}_d(t)) = \boldsymbol{\mathcal{P}}(\boldsymbol{x}_d(t)) = \mathbf{1} - \boldsymbol{\mathcal{Q}}(\boldsymbol{x}_d(t))$. Introducing the $n \times n$ matrix $\boldsymbol{\mathcal{T}}(\boldsymbol{x})$ with components

$$(\boldsymbol{\mathcal{T}}(\boldsymbol{x}))_{il} = \sum_{j=1}^{p}\sum_{k=1}^{n}\frac{\partial}{\partial x_l}\mathcal{B}_{ij}(\boldsymbol{x})\,\mathcal{B}^{+}_{jk}(\boldsymbol{x})\,(\dot{x}_k(t) - R_k(\boldsymbol{x})),\ i,l \in \{1,\ldots,n\}, \tag{2.50}$$

allows a rearrangement of Eq. (2.47) in the form

$$\Delta\dot{\boldsymbol{x}}(t) = \boldsymbol{\mathcal{Q}}(\boldsymbol{x}_d(t))\left(\boldsymbol{R}(\boldsymbol{x}_d(t)) - \dot{\boldsymbol{x}}_d(t)\right) + \left(\nabla\boldsymbol{R}(\boldsymbol{x}_d(t)) + \boldsymbol{\mathcal{T}}(\boldsymbol{x}_d(t))\right)\Delta\boldsymbol{x}(t), \tag{2.51}$$

$$\Delta\boldsymbol{x}(t_0) = \boldsymbol{x}(t_0) - \boldsymbol{x}_d(t_0). \tag{2.52}$$

If $\boldsymbol{x}_d(t)$ is an exactly realizable trajectory, it satisfies the constraint equation (2.38) and the initial condition $\boldsymbol{x}_d(t_0) = \boldsymbol{x}(t_0)$, and Eq. (2.51) simplifies to the linear homogeneous equation for $\Delta\boldsymbol{x}(t)$,

$$\Delta\dot{\boldsymbol{x}}(t) = \left(\nabla\boldsymbol{R}(\boldsymbol{x}_d(t)) + \boldsymbol{\mathcal{T}}(\boldsymbol{x}_d(t))\right)\Delta\boldsymbol{x}(t) \tag{2.53}$$

$$\Delta\boldsymbol{x}(t_0) = \boldsymbol{0}. \tag{2.54}$$

Clearly, Eq. (2.53) has a vanishing solution

$$\Delta\boldsymbol{x}(t) \equiv \boldsymbol{0}. \tag{2.55}$$

In summary, it was proven that if the desired trajectory $\boldsymbol{x}_d(t)$ is an exactly realizable trajectory, then the state trajectory $\boldsymbol{x}(t)$ follows the desired trajectory exactly, i.e., $\boldsymbol{x}(t) = \boldsymbol{x}_d(t)$. Equation (2.12) in Section 2.1.2 proved already the converse: any controlled state trajectory $\boldsymbol{x}(t)$ satisfies the constraint equation. This allows the following conclusion:

The controlled state trajectory $\boldsymbol{x}(t)$ follows the desired trajectory $\boldsymbol{x}_d(t)$ exactly if and only if $\boldsymbol{x}_d(t)$ is an exactly realizable trajectory.

The notion of an exactly realizable trajectory leads to the following interpretation. Not every desired trajectory $\boldsymbol{x}_d(t)$ can be enforced in a specified controlled dynamical system. In general, the desired trajectory $\boldsymbol{x}_d(t)$ is that what you want, but is not what you get. What you get is an exactly realizable trajectory. Because the control signal $\boldsymbol{u}(t)$ consists of only p independent components, it is possible to find at most p one-to-one relations between state components and components of the control signal. Only p components of a state trajectory can be prescribed, while the remaining $n-p$ components are free. The time evolution of these $n-p$ components is given by the constraint equation (2.38). This motivates the name constraint equation. For an arbitrary desired trajectory $\boldsymbol{x}_d(t)$ to be exactly realizable, it has to be constrained by Eq. (2.38). There is still some freedom to choose which state components are actually prescribed, and which have to be determined by the constraint equation. Until further notice, we adopt the canonical view that the part $\boldsymbol{\mathcal{P}}\boldsymbol{x}_d(t)$ is prescribed by the experimenter, while the part $\boldsymbol{\mathcal{Q}}\boldsymbol{x}_d(t)$ of the state vector is fixed by the constraint equation (2.38). This, however, is not the only possibility, and many more choices are possible. The part $\boldsymbol{\mathcal{P}}\boldsymbol{x}_d(t)$ can be seen as an output for the control system which can be enforced exactly if $\boldsymbol{x}_d(t)$ is exactly realizable. Section 2.6 discusses the possibility to realize general desired outputs not necessarily given by $\boldsymbol{\mathcal{P}}\boldsymbol{x}_d(t)$.

Chapter 3 investigates the relation of exactly realizable trajectories with optimal trajectory tracking. The control solution Eq. (2.40) is the solution to a certain optimal trajectory tracking problem. This insight is the starting point in Chapter 4 to obtain analytical approximations to optimal trajectory tracking of desired trajectories which are not exactly realizable.

The necessity to satisfy condition 2 of equal initial condition leaves two possibilities. Either the system is prepared in the initial state $\boldsymbol{x}(t_0) = \boldsymbol{x}_0 = \boldsymbol{x}_d(t_0)$, or the desired trajectory $\boldsymbol{x}_d(t)$ is designed such that it starts from the observed initial system state $\boldsymbol{x}_0$. In any case, the constraint equation (2.38), seen as an ODE for $\boldsymbol{\mathcal{Q}}\boldsymbol{x}_d(t)$, has to be solved with the initial condition $\boldsymbol{\mathcal{Q}}\boldsymbol{x}_d(t_0) = \boldsymbol{\mathcal{Q}}\boldsymbol{x}_0$.

The control solution as given by Eq. (2.40) is an open loop control. As such, it does not guarantee a stable time evolution, and the controlled system does not necessarily follow the realizable trajectory in the presence of perturbations. The linear equation (2.53) encountered during the proof implies statements about the linear stability of realizable trajectories. A non-vanishing initial value $\Delta\boldsymbol{x}(t_0) = \Delta\boldsymbol{x}_0 \neq \boldsymbol{0}$ constitutes a perturbation of the initial conditions of an exactly realizable trajectory. The control approach as proposed here is only a first step. For a specified exactly realizable trajectory, Eq. (2.53) has to be investigated to determine its linear stability properties. If the desired trajectory is linearly unstable, countermeasures in form of an additional feedback control, for example, have to be applied to guarantee a successful control. Stability of exactly realizable trajectories is not discussed in this thesis.

The concept of a realizable trajectory is elucidated with the help of some examples in the following.

Example 2.3: Controlled FHN model with invertible coupling matrix

Consider the controlled FHN model in the form

$$\begin{pmatrix} \dot{x}(t) \\ \dot{y}(t) \end{pmatrix} = \begin{pmatrix} a_0 + a_1 x(t) + a_2 y(t) \\ R(x(t), y(t)) \end{pmatrix} + \begin{pmatrix} u_1(t) \\ u_2(t) \end{pmatrix}. \tag{2.56}$$

The constant coupling matrix $\boldsymbol{\mathcal{B}}$ is identical to the identity,

$$\boldsymbol{\mathcal{B}} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}. \tag{2.57}$$

This system has two state components x, y, and two independent control signals u_1, u_2. The projector $\boldsymbol{\mathcal{P}}$ is simply the identity, $\boldsymbol{\mathcal{P}} = \boldsymbol{1}$, and $\boldsymbol{\mathcal{Q}} = \boldsymbol{0}$ the zero matrix. The constraint equation (2.38) is trivially satisfied. Any desired trajectory $\boldsymbol{x}_d(t)$ is a realizable trajectory as long as initially, the desired trajectory

equals the state trajectory,

$$\boldsymbol{x}_d(t_0) = \boldsymbol{x}(t_0). \tag{2.58}$$

Example 2.4: Mechanical control system in one spatial dimension

The control signal realizing a desired trajectory $\boldsymbol{x}_d(t)$ of a mechanical control system (see Examples 1.1 and 2.1 for more details) is

$$\begin{aligned} u(t) &= \boldsymbol{B}^+(\boldsymbol{x}_d(t))\left(\dot{\boldsymbol{x}}_d(t) - \boldsymbol{R}(\boldsymbol{x}_d(t))\right) \\ &= \frac{1}{B(x_d(t), y_d(t))}\left(\dot{y}_d(t) - R(x_d(t), y_d(t))\right). \end{aligned} \tag{2.59}$$

The non-vanishing component of the constraint equation (2.38) for realizable desired trajectories simply becomes

$$\dot{x}_d(t) = y_d(t). \tag{2.60}$$

With a scalar control signal $u(t)$ only one state component can be controlled. According to our convention, this state component is

$$\boldsymbol{\mathcal{P}}\boldsymbol{x}_d(t) = \begin{pmatrix} 0 \\ y_d(t) \end{pmatrix}. \tag{2.61}$$

The desired velocity over time $y_d(t)$ can be arbitrarily chosen apart from its initial value, which must be identical to the initial state velocity, $y_d(t_0) = y(t_0)$. The corresponding position over time $x_d(t)$ is given by the constraint equation (2.60). Because Eq. (2.60) is a linear differential equation for $x_d(t)$, its solution in terms of the arbitrary velocity $y_d(t)$ is easily obtained as

$$x_d(t) = x_d(t_0) + \int_{t_0}^{t} d\tau y_d(\tau). \tag{2.62}$$

The initial desired position $x_d(t_0)$ has to agree with the initial state position, $x_d(t_0) = x(t_0) = x_0$. With the help of solution (2.62), the control (2.59) can be entirely expressed in terms of the prescribed velocity over time $y_d(t)$ as

$$\begin{aligned} u(t) = &\frac{1}{B\left(x_0 + \int_{t_0}^{t} d\tau y_d(\tau), y_d(t)\right)} \\ &\times \left(\dot{y}_d(t) - R\left(x_0 + \int_{t_0}^{t} d\tau y_d(\tau), y_d(t)\right)\right). \end{aligned} \tag{2.63}$$

Note that an exact solution to the nonlinear controlled state equation as well as to the control signal $\boldsymbol{u}(t)$ is obtained without actually solving any nonlinear equation. The context of a mechanical control system allows the following interpretation of our approach. The constraint equation (2.60) is the definition of the velocity of a point particle, and no external force R or control force Bu can change that definition. With only a single control signal u, position x and velocity y over time cannot be controlled independently from each other.

One might ask if it is possible to control position and velocity independently of each other by introducing an additional control signal. If both control signal act as forces, the controlled mechanical system with state space dimension $n = 2$ and control space dimension $p = 2$ is

$$\begin{pmatrix} \dot{x}(t) \\ \dot{y}(t) \end{pmatrix} = \begin{pmatrix} y(t) \\ R(x(t), y(t)) \end{pmatrix} + \begin{pmatrix} 0 & 0 \\ B_1(x(t), y(t)) & B_2(x(t), y(t)) \end{pmatrix} \begin{pmatrix} u_1(t) \\ u_2(t) \end{pmatrix}, \tag{2.64}$$

such that the 2×2 coupling matrix $\tilde{\boldsymbol{\mathcal{B}}}$ becomes

$$\tilde{\boldsymbol{\mathcal{B}}}(\boldsymbol{x}) = \begin{pmatrix} 0 & 0 \\ B_1(x, y) & B_2(x, y) \end{pmatrix}. \tag{2.65}$$

However, the structure of $\tilde{\boldsymbol{\mathcal{B}}}$ reveals that it violates the condition of full rank. Indeed, for arbitrary functions $B_1 \neq 0$ and $B_2 \neq 0$, the rank of $\tilde{\boldsymbol{\mathcal{B}}}$ is $\operatorname{rank}\left(\tilde{\boldsymbol{\mathcal{B}}}(\boldsymbol{x})\right) = 1$ and therefore smaller than the control space dimension $p = 2$. The computation of the projectors $\boldsymbol{\mathcal{P}}$ and $\boldsymbol{\mathcal{Q}}$ as well as the computation of the control signal u requires the existence of the inverse of $\tilde{\boldsymbol{\mathcal{B}}}^T \tilde{\boldsymbol{\mathcal{B}}}$, which in turn requires $\tilde{\boldsymbol{\mathcal{B}}}$ to have full rank. Our approach cannot be applied to system (2.64) because both control signals u_1 and u_2 act on the same state component. The corresponding control forces are not independent of each other, but can be combined to a single control force $B_1 u_1 + B_2 u_2$.

The constraint equation (2.60) can also be regarded as an algebraic equation for the desired position over time $x_d(t)$. This is an example for a desired output different from the conventional choice Eq. (2.61). Eliminating the position from the control solution Eq. (2.59) yields

$$u(t) = \frac{1}{B(x_d(t), \dot{x}_d(t))} \left(\ddot{x}_d(t) - R(x_d(t), \dot{x}_d(t))\right). \tag{2.66}$$

Equation (2.66) is a special case of the so-called *computed torque formula*. This approach, also known as inverse dynamics, is regularly applied in robotics. For further information, the reader is referred to the literature about robot control (Lewis et al., 1993; de Wit et al., 2012; Angeles, 2013).

Example 2.5: Activator-controlled FHN model

Consider the FHN model from Example 1.2,

$$\begin{pmatrix} \dot{x}(t) \\ \dot{y}(t) \end{pmatrix} = \begin{pmatrix} a_0 + a_1 x(t) + a_2 y(t) \\ R(x(t), y(t)) \end{pmatrix} + \begin{pmatrix} 0 \\ 1 \end{pmatrix} u(t), \tag{2.67}$$

with coupling vector $\boldsymbol{B} = \begin{pmatrix} 0, & 1 \end{pmatrix}^T$ and standard FHN nonlinearity $R(x, y) = y - \frac{1}{3}y^3 - x$. The projectors $\boldsymbol{\mathcal{P}}$ and $\boldsymbol{\mathcal{Q}}$ are readily computed as

$$\boldsymbol{\mathcal{P}} = \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix}, \qquad \boldsymbol{\mathcal{Q}} = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}. \tag{2.68}$$

Being given by $\boldsymbol{\mathcal{P}}\boldsymbol{x}$, the desired activator component over time $y_d(t)$ can be prescribed. For the desired trajectory to be exactly realizable, the desired inhibitor $x_d(t)$ must be determined from the constraint equation

$$\boldsymbol{\mathcal{Q}}\dot{\boldsymbol{x}}_d(t) = \boldsymbol{\mathcal{Q}}\boldsymbol{R}(\boldsymbol{x}_d(t)). \tag{2.69}$$

Writing down only the non-vanishing component of Eq. (2.69) yields

$$\dot{x}_d(t) = a_1 x_d(t) + a_2 y_d(t) + a_0. \tag{2.70}$$

This linear differential equation for $x_d(t)$ with an inhomogeneity is readily solved in terms of the desired activator over time $y_d(t)$,

$$x_d(t) = \frac{a_0}{a_1}\left(e^{a_1(t-t_0)} - 1\right) + e^{a_1(t-t_0)} x_d(t_0) + a_2 \int_{t_0}^{t} d\tau e^{a_1(t-\tau)} y_d(\tau). \tag{2.71}$$

For the desired trajectory $\boldsymbol{x}_d(t)$ to be exactly realizable, it must agree with the initial state $\boldsymbol{x}(t_0)$ of the controlled system. The control is given by

$$u(t) = \dot{y}_d(t) - R(x_d(t), y_d(t)) = \dot{y}_d(t) - y_d(t) + \frac{1}{3}y_d(t)^3 + x_d(t). \tag{2.72}$$

Using the solution Eq. (2.71), the inhibitor variable $x_d(t)$ can be eliminated from the control signal. Consequently, the control can be expressed as a functional of the desired activator variable $y_d(t)$ and the initial desired inhibitor

value $x_d(t_0)$ as

$$u(t) = \dot{y}_d(t) - y_d(t) + \frac{1}{3}y_d(t)^3 + \frac{a_0}{a_1}\left(e^{a_1(t-t_0)} - 1\right) + e^{a_1(t-t_0)}x_d(t_0) + a_2\int_{t_0}^{t} d\tau e^{a_1(t-\tau)}y_d(\tau)\,. \tag{2.73}$$

To evaluate the performance of the control, the control signal Eq. (2.73) is used in Eq. (2.67), and the resulting controlled dynamical system is solved numerically. The numerically obtained state trajectory is compared with the desired reference trajectory. The desired trajectory is chosen as

$$y_d(t) = \sin(20t)\cos(2t)\,, \tag{2.74}$$

and the initial conditions are set to $x(t_0) = x_d(t_0) = y(t_0) = y_d(t_0) = 0$. As expected from Eq. (2.74), the controlled activator $y(t)$ oscillates wildly, see blue solid line in Fig. 2.1 left. The numerically obtained controlled inhibitor $x(t)$ (red dashed line) increases almost linearly. This behavior can easily be understood from the smallness of a_1 in Eq. (2.71) (see Example 1.2 for parameter values). Indeed, in the limit of vanishing a_1,

$$\lim_{a_1\to 0} x_d(t) = a_0(t-t_0) + x_d(t_0) + a_2\int_{t_0}^{t} d\tau y_d(\tau)\,, \tag{2.75}$$

x_d increases linearly in time with coefficient a_0, while the integral term over a periodic function $y_d(t)$ with zero mean vanishes on average. The control signal $u(t)$, being proportional to $\dot{y}(t)$, oscillates as well, see Fig. 2.1 right. Comparing the differences between the controlled state components and its desired counterparts reveals agreement within numerical precision, see Fig. 2.2 left for the activator and Fig. 2.2 right for the inhibitor component, respectively. However, note that the error increases in time, which could indicate a developing instability. The control, being an open loop control, is potentially unstable. It often must be stabilized to guarantee a successful control. Stabilization of exactly realizable trajectories is not discussed in this thesis.

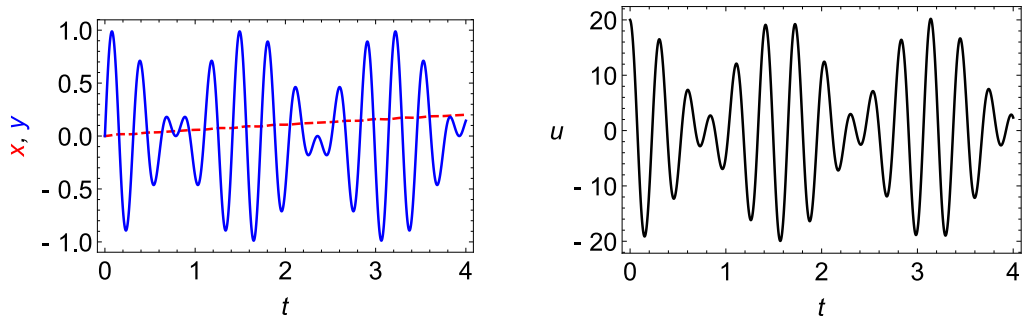

Figure 2.1.: Activator-controlled FHN model driven along an exactly realizable trajectory. The numerically obtained activator y (blue solid line) and inhibitor x (red dashed line) of the controlled system is shown left. The oscillating activator is prescribed according to Eq. (2.74), while the inhibitor cannot be prescribed and is given as the solution to the constraint equation (2.70). The control signal (right) oscillates as well because it is proportional to $\dot{y}_d$.

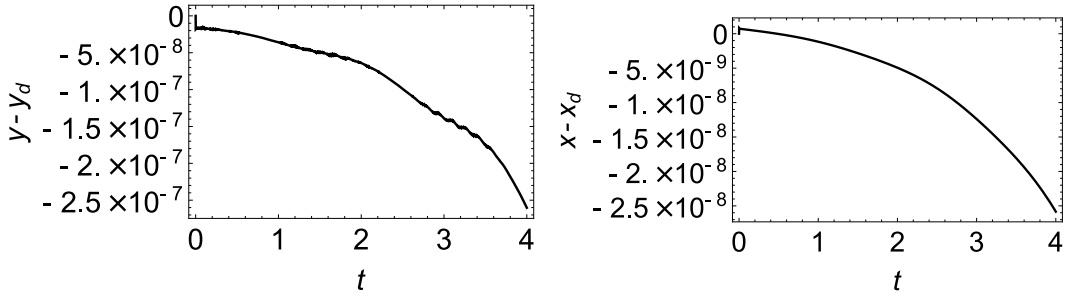

Figure 2.2.: Difference between desired and controlled state components in the activator-controlled FHN model. Plotting the difference between controlled and desired activator $y - y_d$ (left) and controlled and desired inhibitor $x - x_d$ (right) reveals agreement within numerical precision.

Example 2.6: Inhibitor-controlled FHN model

Consider the same model as in Example 2.5 but with a coupling vector $\boldsymbol{B} = \begin{pmatrix} 1, & 0 \end{pmatrix}^T$ corresponding to a control acting on the inhibitor equation (see also Example 1.2).

The projectors are $\boldsymbol{\mathcal{P}} = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}$ and $\boldsymbol{\mathcal{Q}} = \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix}$. The desired inhibitor over time $x_d(t)$ is prescribed while the activator component $y_d(t)$ must be

determined from (the non-vanishing component of) the constraint equation,

$$\dot{y}_d(t) = R(x_d(t), y_d(t)) = y_d(t) - \frac{1}{3} y_d(t)^3 - x_d(t). \tag{2.76}$$

The constraint equation is a nonlinear non-autonomous differential equation for $y_d(t)$. An analytical expression for the solution $y_d(t)$ in terms of the prescribed inhibitor trajectory $x_d(t)$ is not available. Equation (2.76) must be solved numerically.

Figure 2.3 shows the result of a numerical simulation of the controlled system with a desired inhibitor trajectory

$$x_d(t) = 4 \sin(2t), \tag{2.77}$$

and initial conditions $x_d(t_0) = x(t_0) = y_d(t_0) = y(t_0) = 0$. Comparing the desired activator and inhibitor trajectories with the corresponding controlled state trajectories in the bottom panels demonstrates a difference in the range of numerical precision over the whole time interval. Both state components (top left) as well as the control (top right) are oscillating.

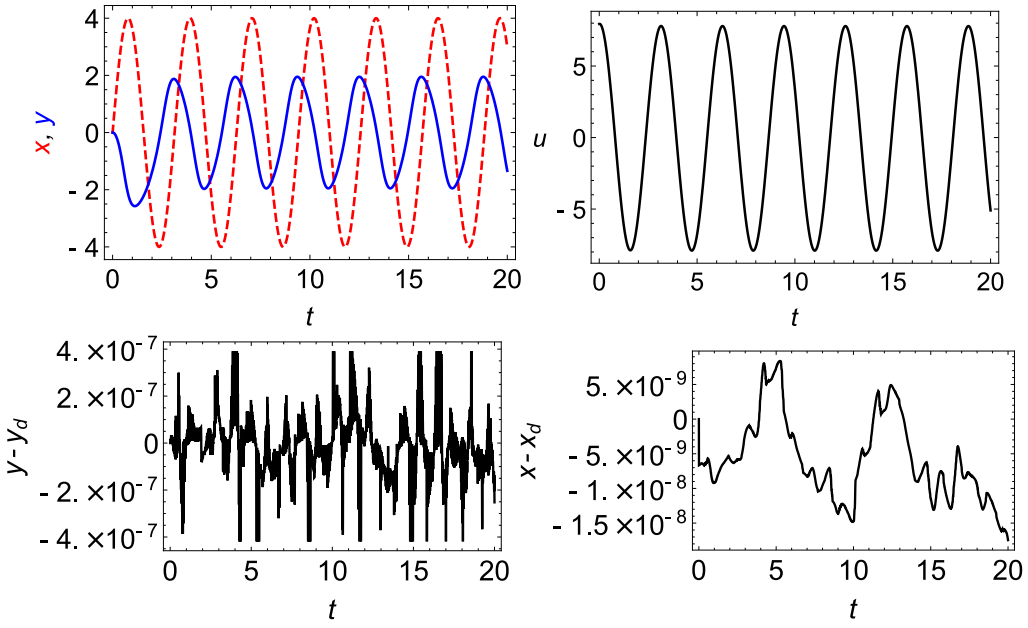

Figure 2.3.: Inhibitor-controlled FHN model driven along an exactly realizable trajectory. The numerically obtained solution of the controlled state is shown top left, and the control signal is shown top right. Comparing desired and controlled activator y_d and y (bottom left) as well as desired and controlled inhibitor x_d and x (bottom left) reveals a difference within numerical precision.

2.3. Linearizing assumption

An uncontrolled dynamical system requires solving

$$\dot{\boldsymbol{x}}(t) = \boldsymbol{R}(\boldsymbol{x}(t)). \tag{2.78}$$

In contrast, control of exactly realizable trajectories requires only the solution of the constraint equation

$$\boldsymbol{\mathcal{Q}}(\boldsymbol{x}(t))(\dot{\boldsymbol{x}}(t) - \boldsymbol{R}(\boldsymbol{x}(t))) = \boldsymbol{0}. \tag{2.79}$$

This opens up the possibility to solve a nonlinear control problem without actually solving any nonlinear equations. If the constraint equation is linear, the entire controlled system can be regarded, in some sense and to some extent, as being linear. Two conditions must be met for Eq. (2.79) to be linear. First of all, the projection matrices $\boldsymbol{\mathcal{P}}(\boldsymbol{x})$ and $\boldsymbol{\mathcal{Q}}(\boldsymbol{x})$ should be independent of the state $\boldsymbol{x}$. This condition can be expressed as

$$\boldsymbol{\mathcal{P}}(\boldsymbol{x}) = \boldsymbol{1} - \boldsymbol{\mathcal{Q}}(\boldsymbol{x}) = \boldsymbol{\mathcal{B}}(\boldsymbol{x})\left(\boldsymbol{\mathcal{B}}^T(\boldsymbol{x})\boldsymbol{\mathcal{B}}(\boldsymbol{x})\right)^{-1}\boldsymbol{\mathcal{B}}^T(\boldsymbol{x}) = \text{const.} \tag{2.80}$$

or

$$\nabla\left(\boldsymbol{\mathcal{B}}(\boldsymbol{x})\left(\boldsymbol{\mathcal{B}}^T(\boldsymbol{x})\boldsymbol{\mathcal{B}}(\boldsymbol{x})\right)^{-1}\boldsymbol{\mathcal{B}}^T(\boldsymbol{x})\right) = \boldsymbol{0}. \tag{2.81}$$

Note that this condition does not imply that the coupling matrix is independent of $\boldsymbol{x}$. Second, the nonlinearity $\boldsymbol{R}(\boldsymbol{x})$ must satisfy

$$\boldsymbol{\mathcal{Q}}\boldsymbol{R}(\boldsymbol{x}) = \boldsymbol{\mathcal{Q}}\boldsymbol{\mathcal{A}}\boldsymbol{x} + \boldsymbol{\mathcal{Q}}\boldsymbol{b}, \tag{2.82}$$

with $n \times n$ matrix $\boldsymbol{\mathcal{A}}$ and n-component vector $\boldsymbol{b}$ independent of the state $\boldsymbol{x}$. Strictly speaking, the projector $\boldsymbol{\mathcal{Q}}$ in front of $\boldsymbol{\mathcal{A}}$ and $\boldsymbol{b}$ is not really necessary. It is placed there to make it clear that $\boldsymbol{\mathcal{A}}$ and $\boldsymbol{b}$ do not contain any parts in the direction of $\boldsymbol{\mathcal{P}}$. Condition Eq. (2.80) combined with condition Eq. (2.82) constitute the *linearizing assumption*. Control systems satisfying the linearizing assumption behave, to a large extent, similar to truly linear control systems. A nonlinear control systems with scalar input $u(t)$ satisfying the linearizing assumption is sometimes said to be in companion form. A system in companion form is trivially feedback linearizable, see the discussion of feedback linearization in Chapter 1 and e.g. (Khalil, 2001).

Condition Eq. (2.82) is a strong assumption. It enforces $n-p$ components of $\boldsymbol{R}(\boldsymbol{x})$ to depend only linearly on the state. However, some important models of nonlinear dynamics satisfy the linearizing assumption. Among these are the mechanical control systems in one spatial dimension, see Examples 1.1 and 2.4, as well as the activator-controlled FHN model discussed in Example 2.5. In both cases, the control signal $\boldsymbol{u}(t)$ acts directly on the nonlinear part of the nonlinearity $\boldsymbol{R}$, such that condition Eq. (2.82) is satisfied. Furthermore, in both cases the coupling matrix $\boldsymbol{\mathcal{B}}(\boldsymbol{x})$ is a coupling vector $\boldsymbol{B}(\boldsymbol{x}) = \begin{pmatrix} 0, & B(x,y) \end{pmatrix}^T$ with only one non-vanishing component. This leads to constant projectors $\boldsymbol{\mathcal{P}}$ and $\boldsymbol{\mathcal{Q}}$, and condition Eq. (2.80) is also satisfied. Another, less obvious example satisfying the linearizing assumption is the controlled SIR model.

Example 2.7: Linearizing assumption satisfied by the controlled SIR model

The controlled state equation for the SIR model was developed in Example 1.3. The nonlinearity $\boldsymbol{R}$ is

$$\boldsymbol{R}\left(\boldsymbol{x}\left(t\right)\right)=\left(-\beta\frac{S\left(t\right)I\left(t\right)}{N},\beta\frac{S\left(t\right)I\left(t\right)}{N}-\gamma I\left(t\right),\gamma I\left(t\right)\right)^{T}, \tag{2.83}$$

while the coupling vector $\boldsymbol{B}$ explicitly depends on the state,

$$\boldsymbol{B}\left(\boldsymbol{x}\left(t\right)\right)=\frac{1}{N}\left(-S\left(t\right)I\left(t\right),S\left(t\right)I\left(t\right),0\right)^{T}. \tag{2.84}$$

However, the projectors

$$\boldsymbol{\mathcal{P}}\left(\boldsymbol{x}\right)=\boldsymbol{\mathcal{P}}=\boldsymbol{B}\left(\boldsymbol{x}\right)\left(\boldsymbol{B}^{T}\left(\boldsymbol{x}\right)\boldsymbol{B}\left(\boldsymbol{x}\right)\right)^{-1}\boldsymbol{B}^{T}\left(\boldsymbol{x}\right)=\frac{1}{2}\begin{pmatrix}1&-1&0\\-1&1&0\\0&0&0\end{pmatrix}, \tag{2.85}$$

$$\boldsymbol{\mathcal{Q}}\left(\boldsymbol{x}\right)=\boldsymbol{\mathcal{Q}}=\frac{1}{2}\begin{pmatrix}1&1&0\\1&1&0\\0&0&2\end{pmatrix}, \tag{2.86}$$

are independent of the state. Furthermore, the model also satisfies the linearizing assumption Eq. (2.80) because

$$\boldsymbol{\mathcal{Q}}\boldsymbol{R}\left(\boldsymbol{x}\right)=\begin{pmatrix}0&-\frac{\gamma}{2}&0\\0&-\frac{\gamma}{2}&0\\0&\gamma&0\end{pmatrix}=\boldsymbol{\mathcal{Q}}\boldsymbol{\mathcal{A}}\boldsymbol{x}. \tag{2.87}$$

The constraint equation is a linear differential equation with three components,

$$\begin{pmatrix}\frac{1}{2}\left(\gamma I_{d}\left(t\right)+\dot{I}_{d}\left(t\right)+\dot{S}_{d}\left(t\right)\right)\\\frac{1}{2}\left(\gamma I_{d}\left(t\right)+\dot{I}_{d}\left(t\right)+\dot{S}_{d}\left(t\right)\right)\\-\gamma I_{d}\left(t\right)+\dot{R}_{d}\left(t\right)\end{pmatrix}=\begin{pmatrix}0\\0\\0\end{pmatrix}, \tag{2.88}$$

of which one is redundant. Note that because the projectors $\boldsymbol{\mathcal{P}}$ and $\boldsymbol{\mathcal{Q}}$ are non-diagonal, the time derivatives of $I_{d}\left(t\right)$ and $S_{d}\left(t\right)$ are mixed in the constraint equation.

2.4. Controllability

A system is called controllable or state controllable if it is possible to achieve a transfer from an initial state $\boldsymbol{x}(t_0) = \boldsymbol{x}_0$ at time $t = t_0$ to a final state $\boldsymbol{x}(t_1) = \boldsymbol{x}_1$ at the terminal time $t = t_1$. Controllability is a condition on the structure of the dynamical system as given by the nonlinearity $\boldsymbol{R}$ and the coupling matrix $\boldsymbol{\mathcal{B}}$. In contrast, for a given control system, trajectory realizability is a condition on the desired trajectory. While for linear control systems controllability is easily expressed in terms of a rank condition, the notion is much more difficult for nonlinear control systems. Section 2.4.1 discusses the Kalman rank condition for the controllability of LTI systems as introduced by Kalman (Kalman, 1959, 1960) in the early sixties. Section 2.4.3 derives a similar rank condition in the context of exactly realizable trajectories. Remarkably, this rank condition also applies to nonlinear systems satisfying the linearizing assumption from Section 2.3.

For general nonlinear systems, the notion of controllability must be refined and it is necessary to distinguish between controllability, accessibility, and reachability. Different and not necessarily equivalent notions of controllability exist, and it is said that there are as many notions of nonlinear controllability as there are researchers in the field. When applied to LTI systems, all of these notions reduce to the Kalman rank condition. Here, no attempt is given to generalize the notion of controllability to nonlinear systems which violate the linearizing assumption. The reader is referred to the literature (Slotine and Li, 1991; Isidori, 1995; Khalil, 2001; Levine, 2009).

2.4.1. Kalman rank condition for LTI systems

Controllability for LTI systems was first introduced by Kalman (Kalman, 1959, 1960). An excellent introduction to linear control systems, including controllability, can be found in (Chen, 1998).

Consider the LTI system with n-dimensional state vector $\boldsymbol{x}(t)$ and p-dimensional control signal $\boldsymbol{u}(t)$,

$$\dot{\boldsymbol{x}}(t) = \boldsymbol{\mathcal{A}}\boldsymbol{x}(t) + \boldsymbol{\mathcal{B}}\boldsymbol{u}(t), \tag{2.89}$$

and initial condition

$$\boldsymbol{x}(t_0) = \boldsymbol{x}_0. \tag{2.90}$$

Here, $\boldsymbol{\mathcal{A}}$ is an $n \times n$ real constant matrix and $\boldsymbol{\mathcal{B}}$ an $n \times p$ real constant matrix. The system Eq. (2.89) is said to be controllable if, for any initial state $\boldsymbol{x}_0$ at the initial time $t = t_0$ and any final state $\boldsymbol{x}_1$ at the terminal time $t = t_1$, there exists an input that transfers $\boldsymbol{x}_0$ to $\boldsymbol{x}_1$. The terminal condition for the state is

$$\boldsymbol{x}(t_1) = \boldsymbol{x}_1. \tag{2.91}$$

The definition of controllability requires only that the input $\boldsymbol{u}(t)$ be capable of moving any state in the state space to any other state in finite time. The state trajectory $\boldsymbol{x}(t)$ traced out in state space is not specified. Kalman showed that this definition of controllability is equivalent to the statement that the $n \times np$ *controllability matrix*

$$\boldsymbol{\mathcal{K}} = \left(\boldsymbol{\mathcal{B}} | \boldsymbol{\mathcal{A}}\boldsymbol{\mathcal{B}} | \boldsymbol{\mathcal{A}}^2\boldsymbol{\mathcal{B}} | \cdots | \boldsymbol{\mathcal{A}}^{n-1}\boldsymbol{\mathcal{B}}\right) \tag{2.92}$$

has rank n, i.e., it satisfies the *Kalman rank condition*

$$\operatorname{rank}(\boldsymbol{\mathcal{K}}) = n. \tag{2.93}$$

Since $n \leq np$, this condition states that $\boldsymbol{\mathcal{K}}$ has full row rank. Equation (2.93) is derived in the following.

2.4.2. Derivation of the Kalman rank condition

The solution $\boldsymbol{x}(t)$ to Eq. (2.89) with initial condition Eq. (2.90) and arbitrary control signal $\boldsymbol{u}(t)$ is

$$\boldsymbol{x}(t) = e^{\boldsymbol{\mathcal{A}}(t-t_0)}\boldsymbol{x}_0 + \int_{t_0}^{t} d\tau e^{\boldsymbol{\mathcal{A}}(t-\tau)}\boldsymbol{\mathcal{B}}\boldsymbol{u}(\tau). \tag{2.94}$$

See also Appendix A.1 for a derivation of the general solution to a forced linear dynamical system. The system is controllable if a control signal $\boldsymbol{u}$ can be found such that the terminal condition (2.91) is satisfied. Evaluating Eq. (2.94) at the terminal time $t = t_1$, multiplying by $e^{-\boldsymbol{\mathcal{A}}(t_1-t_0)}$, rearranging, and expanding the matrix exponential under the integral yields

$$e^{-\boldsymbol{\mathcal{A}}(t_1-t_0)}\boldsymbol{x}_1 - \boldsymbol{x}_0 = \sum_{k=0}^{\infty} \boldsymbol{\mathcal{A}}^k\boldsymbol{\mathcal{B}} \int_{t_0}^{t_1} d\tau \frac{(t_0-\tau)^k}{k!}\boldsymbol{u}(\tau). \tag{2.95}$$

For the system to be controllable, it must in principle be possible to solve for the control signal $\boldsymbol{u}$. As a consequence of the Cayley-Hamilton theorem, the matrix power $\boldsymbol{\mathcal{A}}^i$ for any $n \times n$ matrix with $i \geq n$ can be written as a sum of lower order powers (Fischer, 2013),

$$\boldsymbol{\mathcal{A}}^i = \sum_{k=0}^{n-1} c_{ik}\boldsymbol{\mathcal{A}}^k. \tag{2.96}$$

It follows that the infinite sum in Eq. (2.95) can be rearranged to include only terms with power in $\boldsymbol{\mathcal{A}}$ up to $\boldsymbol{\mathcal{A}}^{n-1}$. The sum on the right hand side (r. h. s.) of Eq.

(2.95) can be simplified as

$$\begin{aligned}
&\sum_{k=0}^{\infty} \boldsymbol{\mathcal{A}}^k \boldsymbol{\mathcal{B}} \int_{t_0}^{t_1} d\tau \frac{(t_0-\tau)^k}{k!} \boldsymbol{u}(\tau) \\
&= \sum_{k=0}^{n-1} \boldsymbol{\mathcal{A}}^k \boldsymbol{\mathcal{B}} \int_{t_0}^{t_1} d\tau \frac{(t_0-\tau)^k}{k!} \boldsymbol{u}(\tau) + \sum_{i=n}^{\infty} \boldsymbol{\mathcal{A}}^i \boldsymbol{\mathcal{B}} \int_{t_0}^{t_1} d\tau \frac{(t_0-\tau)^i}{i!} \boldsymbol{u}(\tau) \\
&= \sum_{k=0}^{n-1} \boldsymbol{\mathcal{A}}^k \boldsymbol{\mathcal{B}} \int_{t_0}^{t_1} d\tau \frac{(t_0-\tau)^k}{k!} \boldsymbol{u}(\tau) + \sum_{i=n}^{\infty} \sum_{k=0}^{n-1} c_{ik} \boldsymbol{\mathcal{A}}^k \boldsymbol{\mathcal{B}} \int_{t_0}^{t_1} d\tau \frac{(t_0-\tau)^i}{i!} \boldsymbol{u}(\tau) \\
&= \sum_{k=0}^{n-1} \boldsymbol{\mathcal{A}}^k \boldsymbol{\mathcal{B}} \int_{t_0}^{t_1} d\tau \left(\frac{(t_0-\tau)^k}{k!} + \sum_{i=n}^{\infty} c_{ik} \frac{(t_0-\tau)^i}{i!} \right) \boldsymbol{u}(\tau) .
\end{aligned} \tag{2.97}$$

It follows that the sum in Eq. (2.95) can be truncated after n terms,

$$e^{-\boldsymbol{\mathcal{A}}(t_1-t_0)} \boldsymbol{x}_1 - \boldsymbol{x}_0 = \sum_{k=0}^{n-1} \boldsymbol{\mathcal{A}}^k \boldsymbol{\mathcal{B}} \boldsymbol{\beta}_k (t_1, t_0) . \tag{2.98}$$

The $\boldsymbol{\beta}_k$ are $p \times 1$ vectors defined as

$$\boldsymbol{\beta}_k (t_1, t_0) = \int_{t_0}^{t_1} d\tau \left(\frac{(t_0-\tau)^k}{k!} + \sum_{i=n}^{\infty} c_{ik} \frac{(t_0-\tau)^i}{i!} \right) \boldsymbol{u}(\tau) , \tag{2.99}$$

which depend on the initial and terminal time t_0 and t_1, respectively. These vectors are functionals of the control $\boldsymbol{u}$ and depend on the matrix $\boldsymbol{\mathcal{A}}$ through the expansion coefficients c_{ik}. Defining the $np \times 1$ vector

$$\boldsymbol{\beta}(t_1, t_0) = \begin{pmatrix} \boldsymbol{\beta}_0 (t_1, t_0) \\ \vdots \\ \boldsymbol{\beta}_{n-1} (t_1, t_0) \end{pmatrix} , \tag{2.100}$$

Eq. (2.98) can be written in terms of $\boldsymbol{\beta}$ and Kalman's controllability $n \times np$ matrix $\boldsymbol{\mathcal{K}}$, Eq. (2.92), as

$$e^{-\boldsymbol{\mathcal{A}}(t_1-t_0)} \boldsymbol{x}_1 - \boldsymbol{x}_0 = \sum_{k=0}^{n-1} \boldsymbol{\mathcal{A}}^k \boldsymbol{\mathcal{B}} \boldsymbol{\beta}_k (t_1, t_0) = \boldsymbol{\mathcal{K}} \boldsymbol{\beta}(t_1, t_0) . \tag{2.101}$$

Equation (2.101) is a linear equation for the vector $\boldsymbol{\beta}(t_1, t_0)$ with inhomogeneity $e^{-\boldsymbol{\mathcal{A}}(t_1-t_0)} \boldsymbol{x}_1 - \boldsymbol{x}_0$. For the system (2.89) to be controllable, every state point $\boldsymbol{x}_1$ must have a corresponding vector $\boldsymbol{\beta}(t_1, t_0)$. In other words, the linear map from

$\boldsymbol{\beta}(t_1, t_0)$ to $\boldsymbol{x}_1$ must be *surjective*. This is the case if and only if the matrix $\boldsymbol{\mathcal{K}}$ has full row rank (Fischer, 2013), i.e.,

$$\operatorname{rank}(\boldsymbol{\mathcal{K}}) = n. \tag{2.102}$$

The Kalman rank condition Eq. (2.102) is a necessary and sufficient condition for the controllability of an LTI system.

A slightly different way to arrive at the same result is to solve (2.101) for the vector $\boldsymbol{\beta}(t_1, t_0)$. A solution in terms of the $n \times n$ matrix $\boldsymbol{\mathcal{K}}\boldsymbol{\mathcal{K}}^T$ is (see also Appendix A.2 how to solve an underdetermined system of equations)

$$\boldsymbol{\beta}(t_1, t_0) = \boldsymbol{\mathcal{K}}^T \left(\boldsymbol{\mathcal{K}}\boldsymbol{\mathcal{K}}^T\right)^{-1} \left(e^{-\boldsymbol{\mathcal{A}}(t_1 - t_0)}\boldsymbol{x}_1 - \boldsymbol{x}_0\right). \tag{2.103}$$

The inverse of $\boldsymbol{\mathcal{K}}\boldsymbol{\mathcal{K}}^T$ does exist only if it has full rank, i.e., $\operatorname{rank}\left(\boldsymbol{\mathcal{K}}\boldsymbol{\mathcal{K}}^T\right) = n$. This is the case if and only if the Kalman rank condition $\operatorname{rank}(\boldsymbol{\mathcal{K}}) = n$ is satisfied.

Controllability has a number of interesting and important consequences. Two examples illustrate the concept and highlight one important consequence.

Example 2.8: Single input diagonal LTI system

We consider an LTI system with state and input matrix

$$\boldsymbol{\mathcal{A}} = \begin{pmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{pmatrix}, \qquad \boldsymbol{\mathcal{B}} = \begin{pmatrix} 1 \\ 1 \end{pmatrix}. \tag{2.104}$$

Kalman's controllability matrix is

$$\boldsymbol{\mathcal{K}} = (\boldsymbol{\mathcal{B}}|\boldsymbol{\mathcal{A}}\boldsymbol{\mathcal{B}}) = \begin{pmatrix} 1 & \lambda_1 \\ 1 & \lambda_2 \end{pmatrix}. \tag{2.105}$$

As long as $\lambda_1 \neq \lambda_2$, $\boldsymbol{\mathcal{K}}$ has rank 2. If $\lambda_1 = \lambda_2$, the second row equals the first row, and $\boldsymbol{\mathcal{K}}$ has rank 1. The system Eq. (2.104) is controllable as long as $\lambda_1 \neq \lambda_2$.

Example 2.9: Two pendulums mounted on a cart

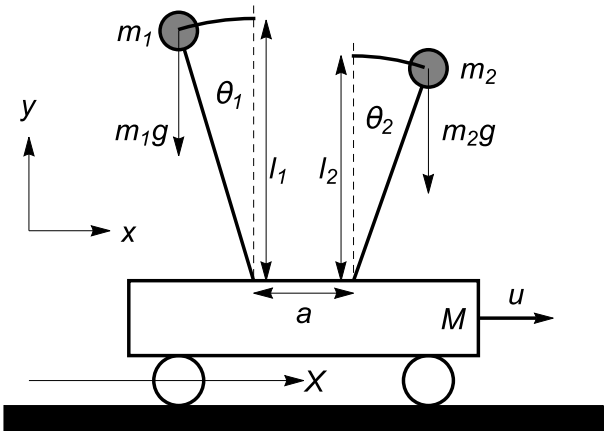

Figure 2.4.: Two inverted pendulums mounted on a cart. The control task is to keep both pendulums in the upright and unstable equilibrium position. The system is controllable as long as the pendulums are not exactly identical, i.e., as long as either their lengths ($l_1 \neq l_2$) or their masses are different ($m_1 \neq m_2$).

Two pendulums mounted on a cart is a mechanical toy model for linear control systems (Chen, 1998). As can be seen in Fig. 2.4, the positions of the masses m_1 and m_2 given by $\begin{pmatrix} x_1, & y_1 \end{pmatrix}^T$ and $\begin{pmatrix} x_2, & y_2 \end{pmatrix}^T$, respectively, are

$$x_1(t) = X(t) - \frac{a}{2} + l_1 \sin(\theta_1(t)), \qquad x_2(t) = X(t) + \frac{a}{2} + l_2 \sin(\theta_2(t)), \tag{2.106}$$

$$y_1(t) = l_1 \cos(\theta_1(t)), \qquad y_2(t) = l_2 \cos(\theta_2(t)). \tag{2.107}$$

The cart can only move in the x direction without any motion in the y-direction. Its position is denoted by $X(t)$. The Lagrangian L equals the difference between kinetic energy T and potential energy V,

$$\begin{aligned} L = T - V = &\frac{1}{2} m_1 \left(\dot{x}_1^2(t) + \dot{y}_1^2(t) \right) + \frac{1}{2} m_2 \left(\dot{x}_2^2(t) + \dot{y}_2^2(t) \right) \\ &+ \frac{1}{2} M \dot{X}^2(t) - m_1 l_1 g \cos(\theta_1(t)) - m_2 l_2 g \cos(\theta_1(t)). \end{aligned} \tag{2.108}$$

The equations of motion are given by the Euler-Lagrange equations

$$\frac{d}{dt}\frac{\partial L}{\partial \dot{\theta}_1} - \frac{\partial L}{\partial \theta_1} = 0, \qquad \frac{d}{dt}\frac{\partial L}{\partial \dot{\theta}_2} - \frac{\partial L}{\partial \theta_2} = 0, \qquad \frac{d}{dt}\frac{\partial L}{\partial \dot{X}} - \frac{\partial L}{\partial X} = u(t). \tag{2.109}$$

The control force $u(t)$ acts on the cart but not on the pendulums. Assuming small angles, $0 \leq |\theta_1(t)| \ll 1$ and $0 \leq |\theta_2(t)| \ll 1$, the equations of motion are linearized around the stationary point. Rewriting the second order differential equations as a controlled dynamical system with $P(t) = M\dot{X}(t)$, $p_1(t) = \dot{\theta}_1$, and $p_2(t) = \dot{\theta}_2(t)$ yields

$$\dot{\boldsymbol{x}}(t) = \boldsymbol{\mathcal{A}}\boldsymbol{x}(t) + \boldsymbol{B}u(t), \tag{2.110}$$

with

$$\boldsymbol{x}(t) = (\theta_1(t), \theta_2(t), X(t), p_1(t), p_2(t), P(t))^T, \tag{2.111}$$

$$\boldsymbol{\mathcal{A}} = \begin{pmatrix} 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & \frac{1}{M} \\ \frac{g(m_1+M)}{l_1 M} & \frac{gm_2}{l_1 M} & 0 & 0 & 0 & 0 \\ \frac{gm_1}{l_2 M} & \frac{g(m_2+M)}{l_2 M} & 0 & 0 & 0 & 0 \\ -gm_1 & -gm_2 & 0 & 0 & 0 & 0 \end{pmatrix}, \tag{2.112}$$

$$\boldsymbol{B} = \left(0, 0, 0, -\frac{1}{l_1 M}, -\frac{1}{l_2 M}, 1\right)^T. \tag{2.113}$$

Kalman's controllability matrix is

$$\begin{aligned} \boldsymbol{\mathcal{K}} &= \left(\boldsymbol{B}|\boldsymbol{\mathcal{A}}\boldsymbol{B}|\boldsymbol{\mathcal{A}}^2\boldsymbol{B}|\cdots|\boldsymbol{\mathcal{A}}^5\boldsymbol{B}\right) \\ &= \begin{pmatrix} 0 & -\frac{1}{l_1 M} & 0 & -\frac{\alpha_3 g}{l_1^2 l_2 M^2} & 0 & -\frac{\beta_1 g^2}{l_1^3 l_2^2 M^3} \\ 0 & -\frac{1}{l_2 M} & 0 & -\frac{\alpha_2 g}{l_1 l_2^2 M^2} & 0 & -\frac{\beta_2 g^2}{l_1^2 l_2^3 M^3} \\ 0 & \frac{1}{M} & 0 & \frac{\alpha_1 g}{l_1 l_2 M^2} & 0 & \frac{\beta_3 g^2}{l_1^2 l_2^2 M^3} \\ -\frac{1}{l_1 M} & 0 & -\frac{\alpha_3 g}{l_1^2 l_2 M^2} & 0 & -\frac{\beta_1 g^2}{l_1^3 l_2^2 M^3} & 0 \\ -\frac{1}{l_2 M} & 0 & -\frac{\alpha_2 g}{l_1 l_2^2 M^2} & 0 & -\frac{\beta_2 g^2}{l_1^2 l_2^3 M^3} & 0 \\ 1 & 0 & \frac{\alpha_1 g}{l_1 l_2 M} & 0 & \frac{\beta_3 g^2}{l_1^2 l_2^2 M^2} & 0 \end{pmatrix} \end{aligned} \tag{2.114}$$

with

$$\alpha_1 = l_2 m_1 + l_1 m_2, \qquad \alpha_2 = l_1(m_2 + M) + l_2 m_1, \tag{2.115}$$

$$\alpha_3 = l_2(m_1 + M) + l_1 m_2, \tag{2.116}$$

and

$$\beta_1 = l_1^2 m_2(m_2 + M) + l_2 l_1 m_2(2m_1 + M) + l_2^2(m_1 + M)^2, \tag{2.117}$$

$$\beta_2 = l_2^2 m_1(m_1 + M) + l_1 l_2 m_1(2m_2 + M) + l_1^2(m_2 + M)^2, \tag{2.118}$$

$$\beta_3 = l_1^2 m_2(m_2 + M) + l_2^2 m_1(m_1 + M) + 2l_2 l_1 m_1 m_2. \tag{2.119}$$

The matrix $\boldsymbol{\mathcal{K}}$ has full row rank,

$$\operatorname{rank}(\boldsymbol{\mathcal{K}}) = 6, \tag{2.120}$$

as long as the pendulums are not identical. Consequently, the system is controllable. A small deviation from the equilibrium position can be counteracted by control. Pendulums with identical mass and lengths, $m_2 = m_1$ and $l_2 = l_1$, respectively, yield

$$\alpha_2 = \alpha_3, \qquad \beta_1 = \beta_2, \tag{2.121}$$

and the first and the second as well as the fourth and the fifth row of the matrix $\boldsymbol{\mathcal{K}}$ become identical. Consequently, the rank of $\boldsymbol{\mathcal{K}}$ changes, and two identical pendulums cannot be controlled.

Both examples show one important consequence of controllability: arbitrary many, parallel connected identical systems cannot be controlled (Kailath, 1980; Chen, 1998). Expressed in a less rigorous language, controllability renders balancing two identical brooms with only a single hand mathematically impossible.

2.4.3. Controllability for systems satisfying the linearizing assumption

Kalman's approach to controllability does not allow a direct generalization to nonlinear systems. Furthermore, nothing is said about the trajectory along which this transfer is achieved. To some extent, these questions can be addressed in the framework of exactly realizable trajectories. Here, a controllability matrix is derived which applies not only to LTI systems but also to nonlinear systems satisfying the linearizing assumption from Section 2.3.

Consider the controlled system

$$\dot{\boldsymbol{x}}(t) = \boldsymbol{R}(\boldsymbol{x}(t)) + \boldsymbol{\mathcal{B}}(\boldsymbol{x}(t))\,\boldsymbol{u}(t) \tag{2.122}$$

together with the linearizing assumption

$$\boldsymbol{\mathcal{Q}}\boldsymbol{R}(\boldsymbol{x}) = \boldsymbol{\mathcal{Q}}\boldsymbol{\mathcal{A}}\boldsymbol{x} + \boldsymbol{\mathcal{Q}}\boldsymbol{b}. \tag{2.123}$$

Equation (2.123) implies a linear constraint equation for an exactly realizable desired trajectory $\boldsymbol{x}_d(t)$,

$$\boldsymbol{\mathcal{Q}}\dot{\boldsymbol{x}}_d(t) = \boldsymbol{\mathcal{Q}}\boldsymbol{\mathcal{A}}\boldsymbol{x}_d(t) + \boldsymbol{\mathcal{Q}}\boldsymbol{b}. \tag{2.124}$$

or, inserting $\mathbf{1} = \boldsymbol{\mathcal{P}} + \boldsymbol{\mathcal{Q}}$ between $\boldsymbol{\mathcal{A}}$ and $\boldsymbol{x}_d$,

$$\boldsymbol{\mathcal{Q}}\dot{\boldsymbol{x}}_d(t) = \boldsymbol{\mathcal{Q}}\boldsymbol{\mathcal{A}}\boldsymbol{\mathcal{Q}}\boldsymbol{x}_d(t) + \boldsymbol{\mathcal{Q}}\boldsymbol{\mathcal{A}}\boldsymbol{\mathcal{P}}\boldsymbol{x}_d(t) + \boldsymbol{\mathcal{Q}}\boldsymbol{b}. \tag{2.125}$$

From now on, the parts $\boldsymbol{\mathcal{P}}\boldsymbol{x}_d(t)$ and $\boldsymbol{\mathcal{Q}}\boldsymbol{x}_d(t)$ are considered as independent state components. The part $\boldsymbol{\mathcal{P}}\boldsymbol{x}_d(t)$ is prescribed by the experimenter while the part $\boldsymbol{\mathcal{Q}}\boldsymbol{x}_d(t)$ is governed by Eq. (2.124). Equation (2.125) is a linear dynamical system for the variable $\boldsymbol{\mathcal{Q}}\boldsymbol{x}_d(t)$ with inhomogeneity $\boldsymbol{\mathcal{Q}}\boldsymbol{\mathcal{A}}\boldsymbol{\mathcal{P}}\boldsymbol{x}_d(t)+\boldsymbol{\mathcal{Q}}\boldsymbol{b}$. Achieving a transfer from the initial state $\boldsymbol{x}_0$ to the finite state $\boldsymbol{x}_1$ means the realizable trajectory $\boldsymbol{x}_d(t)$ has to satisfy

$$\boldsymbol{x}_d(t_0) = \boldsymbol{x}_0, \tag{2.126}$$
$$\boldsymbol{x}_d(t_1) = \boldsymbol{x}_1. \tag{2.127}$$

Consequently, the prescribed part $\boldsymbol{\mathcal{P}}\boldsymbol{x}_d(t)$ satisfies

$$\boldsymbol{\mathcal{P}}\boldsymbol{x}_d(t_0) = \boldsymbol{\mathcal{P}}\boldsymbol{x}_0, \qquad \boldsymbol{\mathcal{P}}\boldsymbol{x}_d(t_1) = \boldsymbol{\mathcal{P}}\boldsymbol{x}_1, \tag{2.128}$$

while the part $\boldsymbol{\mathcal{Q}}\boldsymbol{x}_d(t)$ satisfies

$$\boldsymbol{\mathcal{Q}}\boldsymbol{x}_d(t_0) = \boldsymbol{\mathcal{Q}}\boldsymbol{x}_0, \qquad \boldsymbol{\mathcal{Q}}\boldsymbol{x}_d(t_1) = \boldsymbol{\mathcal{Q}}\boldsymbol{x}_1. \tag{2.129}$$

Being a linear equation, the solution $\boldsymbol{\mathcal{Q}}\boldsymbol{x}_d(t)$ to the constraint equation (2.125) can be expressed as a functional of $\boldsymbol{\mathcal{P}}\boldsymbol{x}_d(t)$,

$$\begin{aligned}\boldsymbol{\mathcal{Q}}\boldsymbol{x}_d(t) = {} & \exp\left(\boldsymbol{\mathcal{Q}}\boldsymbol{\mathcal{A}}\boldsymbol{\mathcal{Q}}\,(t-t_0)\right)\boldsymbol{\mathcal{Q}}\boldsymbol{x}_0 \\ & + \int_{t_0}^{t} d\tau \exp\left(\boldsymbol{\mathcal{Q}}\boldsymbol{\mathcal{A}}\boldsymbol{\mathcal{Q}}\,(t-\tau)\right)\boldsymbol{\mathcal{Q}}\left(\boldsymbol{\mathcal{A}}\boldsymbol{\mathcal{P}}\boldsymbol{x}_d(\tau)+\boldsymbol{b}\right).\end{aligned} \tag{2.130}$$

See also Appendix A.1 for a derivation of the general solution to a forced linear dynamical system. The solution Eq. (2.130) satisfies the initial condition given by Eq. (2.129). Now, all initial and terminal conditions except $\boldsymbol{\mathcal{Q}}\boldsymbol{x}_d(t_1) = \boldsymbol{\mathcal{Q}}\boldsymbol{x}_1$ are satisfied. Enforcing this remaining terminal condition onto the solution Eq. (2.130) yields

$$\begin{aligned}\boldsymbol{\mathcal{Q}}\boldsymbol{x}_1 &= \boldsymbol{\mathcal{Q}}\boldsymbol{x}_d(t_1) \\ &= \exp\left(\boldsymbol{\mathcal{Q}}\boldsymbol{\mathcal{A}}\boldsymbol{\mathcal{Q}}\,(t_1-t_0)\right)\boldsymbol{\mathcal{Q}}\boldsymbol{x}_0 + \int_{t_0}^{t_1} d\tau \exp\left(\boldsymbol{\mathcal{Q}}\boldsymbol{\mathcal{A}}\boldsymbol{\mathcal{Q}}\,(t_1-\tau)\right)\boldsymbol{\mathcal{Q}}\left(\boldsymbol{\mathcal{A}}\boldsymbol{\mathcal{P}}\boldsymbol{x}_d(\tau)+\boldsymbol{b}\right).\end{aligned} \tag{2.131}$$

This is actually a condition for the part $\boldsymbol{\mathcal{P}}\boldsymbol{x}_d(t)$. Therefore, the transfer from $\boldsymbol{x}_0$ to $\boldsymbol{x}_1$ is achieved as long as the part $\boldsymbol{\mathcal{P}}\boldsymbol{x}_d$ satisfies Eqs. (2.128) and (2.131). In between t_0 and t_1, the part $\boldsymbol{\mathcal{P}}\boldsymbol{x}_d(t)$ of the realizable trajectory can be freely chosen by the experimenter. A system is controllable if at least one exactly realizable trajectory $\boldsymbol{x}_d(t)$ can be found such that the constraints Eqs. (2.128) and (2.131) are satisfied.

Analogously to the derivation of the Kalman rank condition in Section 2.4.2, one can ask for the conditions on the state matrices $\boldsymbol{\mathcal{A}}$ and projectors $\boldsymbol{\mathcal{P}}$ and $\boldsymbol{\mathcal{Q}}$ such

that the constraint Eq. (2.131) can be satisfied. Equation (2.131) is rearranged as

$$\exp\left(-\boldsymbol{\mathcal{Q}}\boldsymbol{\mathcal{A}}\boldsymbol{\mathcal{Q}}\left(t_1-t_0\right)\right)\boldsymbol{\mathcal{Q}}\boldsymbol{x}_1-\boldsymbol{\mathcal{Q}}\boldsymbol{x}_0-\int_{t_0}^{t_1}d\tau\exp\left(\boldsymbol{\mathcal{Q}}\boldsymbol{\mathcal{A}}\boldsymbol{\mathcal{Q}}\left(t_0-\tau\right)\right)\boldsymbol{\mathcal{Q}}\boldsymbol{b}$$

$$=\int_{t_0}^{t_1}d\tau\exp\left(\boldsymbol{\mathcal{Q}}\boldsymbol{\mathcal{A}}\boldsymbol{\mathcal{Q}}\left(t_0-\tau\right)\right)\boldsymbol{\mathcal{Q}}\boldsymbol{\mathcal{A}}\boldsymbol{\mathcal{P}}\boldsymbol{x}_d\left(\tau\right), \tag{2.132}$$

and an argument equivalent to the derivation of the Kalman rank condition (2.102) is applied. Due to the Cayley-Hamilton theorem, any power of matrices with $i \geq n$ can be expanded in terms of lower order matrix powers as

$$\left(\boldsymbol{\mathcal{Q}}\boldsymbol{\mathcal{A}}\boldsymbol{\mathcal{Q}}\right)^i=\sum_{k=0}^{n-1}d_{ik}\left(\boldsymbol{\mathcal{Q}}\boldsymbol{\mathcal{A}}\boldsymbol{\mathcal{Q}}\right)^k. \tag{2.133}$$

The r. h. s. of Eq. (2.132) can be simplified as

$$\begin{aligned}
&\int_{t_0}^{t_1}d\tau\exp\left(\boldsymbol{\mathcal{Q}}\boldsymbol{\mathcal{A}}\boldsymbol{\mathcal{Q}}\left(t_0-\tau\right)\right)\boldsymbol{\mathcal{Q}}\boldsymbol{\mathcal{A}}\boldsymbol{\mathcal{P}}\boldsymbol{x}_d\left(\tau\right)\\
&=\sum_{k=0}^{\infty}\left(\boldsymbol{\mathcal{Q}}\boldsymbol{\mathcal{A}}\boldsymbol{\mathcal{Q}}\right)^k\boldsymbol{\mathcal{Q}}\boldsymbol{\mathcal{A}}\boldsymbol{\mathcal{P}}\int_{t_0}^{t_1}d\tau\frac{\left(t_0-\tau\right)^k}{k!}\boldsymbol{\mathcal{P}}\boldsymbol{x}_d\left(\tau\right)\\
&=\sum_{k=0}^{n-1}\left(\boldsymbol{\mathcal{Q}}\boldsymbol{\mathcal{A}}\boldsymbol{\mathcal{Q}}\right)^k\boldsymbol{\mathcal{Q}}\boldsymbol{\mathcal{A}}\boldsymbol{\mathcal{P}}\int_{t_0}^{t_1}d\tau\frac{\left(t_0-\tau\right)^k}{k!}\boldsymbol{\mathcal{P}}\boldsymbol{x}_d\left(\tau\right)\\
&\quad+\sum_{i=n}^{\infty}\left(\boldsymbol{\mathcal{Q}}\boldsymbol{\mathcal{A}}\boldsymbol{\mathcal{Q}}\right)^i\boldsymbol{\mathcal{Q}}\boldsymbol{\mathcal{A}}\boldsymbol{\mathcal{P}}\int_{t_0}^{t_1}d\tau\frac{\left(t_0-\tau\right)^i}{i!}\boldsymbol{\mathcal{P}}\boldsymbol{x}_d\left(\tau\right)\\
&=\sum_{k=0}^{n-1}\left(\boldsymbol{\mathcal{Q}}\boldsymbol{\mathcal{A}}\boldsymbol{\mathcal{Q}}\right)^k\boldsymbol{\mathcal{Q}}\boldsymbol{\mathcal{A}}\boldsymbol{\mathcal{P}}\int_{t_0}^{t_1}d\tau\left(\frac{\left(t_0-\tau\right)^k}{k!}+\sum_{i=n}^{\infty}d_{ik}\frac{\left(t_0-\tau\right)^i}{i!}\right)\boldsymbol{\mathcal{P}}\boldsymbol{x}_d\left(\tau\right),
\end{aligned} \tag{2.134}$$

such that Eq. (2.132) becomes a truncated sum

$$\exp\left(-\boldsymbol{\mathcal{Q}}\boldsymbol{\mathcal{A}}\boldsymbol{\mathcal{Q}}\left(t_1-t_0\right)\right)\boldsymbol{\mathcal{Q}}\boldsymbol{x}_1-\boldsymbol{\mathcal{Q}}\boldsymbol{x}_0-\int_{t_0}^{t_1}d\tau\exp\left(\boldsymbol{\mathcal{Q}}\boldsymbol{\mathcal{A}}\boldsymbol{\mathcal{Q}}\left(t_0-\tau\right)\right)\boldsymbol{\mathcal{Q}}\boldsymbol{b}$$

$$=\sum_{k=0}^{n-1}\left(\boldsymbol{\mathcal{Q}}\boldsymbol{\mathcal{A}}\boldsymbol{\mathcal{Q}}\right)^k\boldsymbol{\mathcal{Q}}\boldsymbol{\mathcal{A}}\boldsymbol{\mathcal{P}}\boldsymbol{\alpha}_k\left(t_1,t_0\right). \tag{2.135}$$

Define the $n \times 1$ vectors

$$\boldsymbol{\alpha}_k\left(t_1,t_0\right)=\int_{t_0}^{t_1}d\tau\left(\frac{\left(t_0-\tau\right)^k}{k!}+\sum_{i=n}^{\infty}d_{ik}\frac{\left(t_0-\tau\right)^i}{i!}\right)\boldsymbol{\mathcal{P}}\boldsymbol{x}_d\left(\tau\right). \tag{2.136}$$

The right hand side of Eq. (2.135) can be written with the help of the $n^2 \times 1$ vector

$$\boldsymbol{\alpha}(t_1, t_0) = \begin{pmatrix} \boldsymbol{\alpha}_0(t_1, t_0) \\ \boldsymbol{\alpha}_1(t_1, t_0) \\ \vdots \\ \boldsymbol{\alpha}_{n-1}(t_1, t_0) \end{pmatrix} \tag{2.137}$$

as

$$\begin{aligned} &\exp\left(-\boldsymbol{\mathcal{Q}}\boldsymbol{\mathcal{A}}\boldsymbol{\mathcal{Q}}(t_1 - t_0)\right)\boldsymbol{\mathcal{Q}}\boldsymbol{x}_1 - \boldsymbol{\mathcal{Q}}\boldsymbol{x}_0 \\ &\quad - \int_{t_0}^{t_1} d\tau \exp\left(\boldsymbol{\mathcal{Q}}\boldsymbol{\mathcal{A}}\boldsymbol{\mathcal{Q}}(t_0 - \tau)\right)\boldsymbol{\mathcal{Q}}\boldsymbol{b} = \tilde{\boldsymbol{\mathcal{K}}}\boldsymbol{\alpha}(t_1, t_0). \end{aligned} \tag{2.138}$$

The $n \times n^2$ *controllability matrix* $\tilde{\boldsymbol{\mathcal{K}}}$ is defined by

$$\tilde{\boldsymbol{\mathcal{K}}} = \left(\boldsymbol{\mathcal{Q}}\boldsymbol{\mathcal{A}}\boldsymbol{\mathcal{P}} | \boldsymbol{\mathcal{Q}}\boldsymbol{\mathcal{A}}\boldsymbol{\mathcal{Q}}\boldsymbol{\mathcal{A}}\boldsymbol{\mathcal{P}} | \cdots | \left(\boldsymbol{\mathcal{Q}}\boldsymbol{\mathcal{A}}\boldsymbol{\mathcal{Q}}\right)^{n-1}\boldsymbol{\mathcal{Q}}\boldsymbol{\mathcal{A}}\boldsymbol{\mathcal{P}}\right). \tag{2.139}$$

The left hand side of Eq. (2.138) can be any point in $\boldsymbol{\mathcal{Q}}\mathbb{R}^n = \mathbb{R}^{n-p}$. The mapping is surjective, i.e., every element on the left hand side has a corresponding element on the right hand side, if $\tilde{\boldsymbol{\mathcal{K}}}$ has full rank $n - p$. Therefore, the nonlinear affine control system Eq. (2.122) satisfying the linearizing assumption Eq. (2.123) is controllable if

$$\operatorname{rank}\left(\tilde{\boldsymbol{\mathcal{K}}}\right) = n - p. \tag{2.140}$$

Example 2.10: Single input diagonal LTI system

Consider the LTI system from Example 2.8. The two parts of $\boldsymbol{\mathcal{A}}$ necessary for the computation of the controllability matrix $\tilde{\boldsymbol{\mathcal{K}}}$ are

$$\boldsymbol{\mathcal{Q}}\boldsymbol{\mathcal{A}}\boldsymbol{\mathcal{P}} = \frac{1}{4}\begin{pmatrix} \lambda_1 - \lambda_2 & \lambda_1 - \lambda_2 \\ \lambda_2 - \lambda_1 & \lambda_2 - \lambda_1 \end{pmatrix}, \quad \boldsymbol{\mathcal{Q}}\boldsymbol{\mathcal{A}}\boldsymbol{\mathcal{Q}} = \frac{1}{4}\begin{pmatrix} \lambda_1 + \lambda_2 & \lambda_1 - \lambda_2 \\ \lambda_2 - \lambda_1 & \lambda_1 + \lambda_2 \end{pmatrix}. \tag{2.141}$$

The controllability matrix is

$$\begin{aligned} \tilde{\boldsymbol{\mathcal{K}}} &= \left(\boldsymbol{\mathcal{Q}}\boldsymbol{\mathcal{A}}\boldsymbol{\mathcal{P}} | \boldsymbol{\mathcal{Q}}\boldsymbol{\mathcal{A}}\boldsymbol{\mathcal{Q}}\boldsymbol{\mathcal{A}}\boldsymbol{\mathcal{P}}\right) \\ &= \frac{1}{4}\begin{pmatrix} (\lambda_1 - \lambda_2) & (\lambda_1 - \lambda_2) & \frac{1}{2}(\lambda_1^2 - \lambda_2^2) & \frac{1}{2}(\lambda_1^2 - \lambda_2^2) \\ (\lambda_2 - \lambda_1) & (\lambda_2 - \lambda_1) & \frac{1}{2}(\lambda_2^2 - \lambda_1^2) & \frac{1}{2}(\lambda_2^2 - \lambda_1^2) \end{pmatrix}. \end{aligned} \tag{2.142}$$

The upper row of $\tilde{\boldsymbol{\mathcal{K}}}$ equals the lower row times -1, i.e., the rows are linearly dependent and so $\tilde{\boldsymbol{\mathcal{K}}}$ has rank

$$\operatorname{rank}\left(\tilde{\boldsymbol{\mathcal{K}}}\right) = 1. \tag{2.143}$$

If $\lambda_1 = \lambda_2$, all entries of $\tilde{\boldsymbol{\mathcal{K}}}$ vanish and then $\tilde{\boldsymbol{\mathcal{K}}}$ has zero rank. The system Eq. (2.104) is controllable as long as $\lambda_1 \neq \lambda_2$.

Example 2.11: Controllability of the activator-controlled FHN model

Controllability in form of a rank condition can be discussed for all models of the form

$$\begin{pmatrix} \dot{x}(t) \\ \dot{y}(t) \end{pmatrix} = \begin{pmatrix} a_0 + a_1 x(t) + a_2 y(t) \\ R(x(t), y(t)) \end{pmatrix} + \begin{pmatrix} 0 \\ B(x(t), y(t)) \end{pmatrix} u(t). \quad (2.144)$$

A prominent example is the activator-controlled FHN model. The $\boldsymbol{\mathcal{Q}}$ part of the nonlinearity $\boldsymbol{R}$ is actually a linear function of the state $\boldsymbol{x}$,

$$\boldsymbol{\mathcal{Q}R}(\boldsymbol{x}(t)) = \begin{pmatrix} a_1 x(t) + a_2 y(t) \\ 0 \end{pmatrix} + \begin{pmatrix} a_0 \\ 0 \end{pmatrix} = \boldsymbol{\mathcal{Q}\mathcal{A}x}(t) + \boldsymbol{\mathcal{Q}b}, \quad (2.145)$$

i.e., this model satisfies the linearizing assumption with the matrix $\boldsymbol{\mathcal{A}}$ and vector $\boldsymbol{b}$ defined by

$$\boldsymbol{\mathcal{Q}\mathcal{A}} = \begin{pmatrix} a_1 & a_2 \\ 0 & 0 \end{pmatrix}, \qquad \boldsymbol{\mathcal{Q}b} = \begin{pmatrix} a_0 \\ 0 \end{pmatrix}. \quad (2.146)$$

The controllability matrix $\tilde{\boldsymbol{\mathcal{K}}}$ is

$$\tilde{\boldsymbol{\mathcal{K}}} = (\boldsymbol{\mathcal{Q}\mathcal{A}\mathcal{P}} | \boldsymbol{\mathcal{Q}\mathcal{A}\mathcal{Q}\mathcal{A}\mathcal{P}}) = \begin{pmatrix} 0 & a_2 & 0 & a_1 a_2 \\ 0 & 0 & 0 & 0 \end{pmatrix}, \quad (2.147)$$

and, for $a_2 \neq 0$, $\tilde{\boldsymbol{\mathcal{K}}}$ has rank

$$\operatorname{rank}\left(\tilde{\boldsymbol{\mathcal{K}}}\right) = 1 = n - p. \quad (2.148)$$

The activator-controlled FHN model is controllable as long as $a_2 \neq 0$, i.e., as long as the equation for the inhibitor x also depends on the activator y. The control directly affects the activator y. If $a_2 = 0$ in Eq. (2.144), the inhibitor evolves decoupled from the activator, and therefore cannot be affected by control.

Example 2.12: Controllability of the controlled SIR model

With the help of the projectors $\boldsymbol{\mathcal{P}}$ and $\boldsymbol{\mathcal{Q}}$ computed in Example 2.7, the controllability matrix is obtained as

$$\begin{aligned}\tilde{\boldsymbol{\mathcal{K}}} &= (\boldsymbol{\mathcal{QAP}}|\boldsymbol{\mathcal{QAQAP}}|\boldsymbol{\mathcal{QAQAQAP}}) \\ &= \begin{pmatrix} \frac{\gamma}{4} & -\frac{\gamma}{4} & 0 & -\frac{\gamma^2}{8} & \frac{\gamma^2}{8} & 0 & \frac{\gamma^3}{16} & -\frac{\gamma^3}{16} & 0 \\ \frac{\gamma}{4} & -\frac{\gamma}{4} & 0 & -\frac{\gamma^2}{8} & \frac{\gamma^2}{8} & 0 & \frac{\gamma^3}{16} & -\frac{\gamma^3}{16} & 0 \\ -\frac{\gamma}{2} & \frac{\gamma}{2} & 0 & \frac{\gamma^2}{4} & -\frac{\gamma^2}{4} & 0 & -\frac{\gamma^3}{8} & \frac{\gamma^3}{8} & 0 \end{pmatrix}. \end{aligned} \tag{2.149}$$

As long as $\gamma \neq 0$, the rank of $\tilde{\boldsymbol{\mathcal{K}}}$ is

$$\operatorname{rank}\left(\tilde{\boldsymbol{\mathcal{K}}}\right) = 1 < n - p = 2. \tag{2.150}$$

Thus, the rank of $\tilde{\boldsymbol{\mathcal{K}}}$ is smaller than $n-p$, and consequently the SIR model is *not* controllable. It is impossible to find a control to reach every final state $\boldsymbol{x}_1$ from every other initial state $\boldsymbol{x}_0$. Intuitively, the reason is simple to understand. The controlled SIR model satisfies a conservation law, see Example 1.3. Independent of the actual time dependence of the control signal $u(t)$, the total number N of individuals is conserved,

$$S(t) + I(t) + R(t) = N. \tag{2.151}$$

The value of N is prescribed by the initial condition $\boldsymbol{x}(t_0) = \boldsymbol{x}_0$. For all times, the dynamics of the controlled SIR model is restricted to a two-dimensional surface embedded in the three-dimensional state space. Hence, the system's state vector can only reach points lying on this surface, and no control can force the system to leave it.

2.4.4. Discussion

A controllability matrix $\tilde{\boldsymbol{\mathcal{K}}}$, Eq. (2.139), is derived in the framework of exactly realizable trajectories. If $\tilde{\boldsymbol{\mathcal{K}}}$ satisfies the rank condition $\operatorname{rank}\left(\tilde{\boldsymbol{\mathcal{K}}}\right) = n - p$, the system is controllable. At least one control signal exists which achieves a transfer from an arbitrary initial state $\boldsymbol{x}(t_0) = \boldsymbol{x}_0$ to an arbitrary final state $\boldsymbol{x}(t_1) = \boldsymbol{x}_1$ within the finite time interval $t_1 - t_0$.

The controllability matrix $\tilde{\boldsymbol{\mathcal{K}}}$ can be computed for all LTI system. We expect that the rank condition for controllability, Eq. (2.140) is fully equivalent to Kalman's rank condition, Eq. (2.102). If the system is controllable in terms of $\tilde{\boldsymbol{\mathcal{K}}}$, it is also controllable in terms of $\boldsymbol{\mathcal{K}}$, and vice versa. The advantage of controllability in terms of $\tilde{\boldsymbol{\mathcal{K}}}$ is its applicability to a certain class of nonlinear systems. For affine dynamical systems satisfying the linearizing assumption (2.82), the rank condition for $\tilde{\boldsymbol{\mathcal{K}}}$

remains a valid check for controllability. This class encompasses a number of simple nonlinear models which are of interest to physicists. In particular, it is proven that all mechanical control systems in one spatial dimension are controllable, see Example 2.11. Other systems satisfying the linearizing assumption are the controlled SIR model, Example 2.12, and the activator-controlled FHN model, see Example 2.11. Controllability as proposed here cannot be applied to the inhibitor-controlled FHN model because the corresponding constraint equation is nonlinear. Checking its controllability requires a notion of nonlinear controllability for general nonlinear systems. Nonlinear controllability cannot be defined in form of a simple rank condition for a controllability matrix but demands more difficult concepts.

Exactly realizable trajectories allow a characterization of the entirety of state trajectories along which a state transfer can be achieved. Any desired trajectory which satisfies the constraint equation

$$\boldsymbol{\mathcal{Q}}\left(\dot{\boldsymbol{x}}_d\left(t\right)-\boldsymbol{\mathcal{A}}\boldsymbol{x}_d\left(t\right)-\boldsymbol{b}\right)=\boldsymbol{0}, \tag{2.152}$$

and the initial and terminal conditions

$$\boldsymbol{x}_d\left(t_0\right)=\boldsymbol{x}_0, \qquad\qquad \boldsymbol{x}_d\left(t_1\right)=\boldsymbol{x}_1 \tag{2.153}$$

does the job. For example, a second order differential equation for $\boldsymbol{x}_d\left(t\right)$ can accommodate both initial and terminal conditions Eqs. (2.153). A successful transfer from $\boldsymbol{x}_0$ to $\boldsymbol{x}_1$ is achieved if $\boldsymbol{x}_d\left(t\right)$ additionally satisfies the constraint equation (2.152). The control signal is given by

$$\boldsymbol{u}\left(t\right)=\boldsymbol{\mathcal{B}}^{+}\left(\dot{\boldsymbol{x}}_d\left(t\right)-\boldsymbol{\mathcal{A}}\boldsymbol{x}_d\left(t\right)-\boldsymbol{b}\right). \tag{2.154}$$

Equation (2.154) can be used to obtain an expression for the control which depends only on the part $\boldsymbol{\mathcal{P}}\boldsymbol{x}_d\left(t\right)$. According to equation (2.130), the solution for $\boldsymbol{\mathcal{Q}}\boldsymbol{x}_d\left(t\right)$ can be expressed in terms of a functional of $\boldsymbol{\mathcal{P}}\boldsymbol{x}_d\left(t\right)$,

$$\begin{aligned}\boldsymbol{\mathcal{Q}}\boldsymbol{x}_d\left[\boldsymbol{\mathcal{P}}\boldsymbol{x}_d\left(t\right)\right]=&\exp\left(\boldsymbol{\mathcal{Q}}\boldsymbol{\mathcal{A}}\boldsymbol{\mathcal{Q}}\left(t-t_0\right)\right)\boldsymbol{\mathcal{Q}}\boldsymbol{x}_0\\&+\int_{t_0}^{t}d\tau\exp\left(\boldsymbol{\mathcal{Q}}\boldsymbol{\mathcal{A}}\boldsymbol{\mathcal{Q}}\left(t-\tau\right)\right)\boldsymbol{\mathcal{Q}}\left(\boldsymbol{\mathcal{A}}\boldsymbol{\mathcal{P}}\boldsymbol{x}_d\left(\tau\right)+\boldsymbol{b}\right).\end{aligned} \tag{2.155}$$

Consequently, Eq. (2.154) becomes a functional of $\boldsymbol{\mathcal{P}}\boldsymbol{x}_d\left(t\right)$ as well,

$$\begin{aligned}\boldsymbol{u}\left[\boldsymbol{\mathcal{P}}\boldsymbol{x}_d\left(t\right)\right]=&\boldsymbol{\mathcal{B}}^{+}\left(\boldsymbol{\mathcal{P}}\dot{\boldsymbol{x}}_d\left(t\right)-\boldsymbol{\mathcal{P}}\boldsymbol{\mathcal{A}}\boldsymbol{\mathcal{P}}\boldsymbol{x}_d\left(t\right)-\boldsymbol{\mathcal{P}}\boldsymbol{\mathcal{A}}\exp\left(\boldsymbol{\mathcal{Q}}\boldsymbol{\mathcal{A}}\boldsymbol{\mathcal{Q}}\left(t-t_0\right)\right)\boldsymbol{\mathcal{Q}}\boldsymbol{x}_0-\boldsymbol{\mathcal{P}}\boldsymbol{b}\right)\\&-\boldsymbol{\mathcal{B}}^{+}\boldsymbol{\mathcal{A}}\int_{t_0}^{t}d\tau\exp\left(\boldsymbol{\mathcal{Q}}\boldsymbol{\mathcal{A}}\boldsymbol{\mathcal{Q}}\left(t-\tau\right)\right)\boldsymbol{\mathcal{Q}}\left(\boldsymbol{\mathcal{A}}\boldsymbol{\mathcal{P}}\boldsymbol{x}_d\left(\tau\right)+\boldsymbol{b}\right).\end{aligned} \tag{2.156}$$

Thus, any reference to $\boldsymbol{\mathcal{Q}}\boldsymbol{x}_d\left(t\right)$ except for the initial condition $\boldsymbol{\mathcal{Q}}\boldsymbol{x}_0$ is eliminated from the expression for the control signal. The control signal is entirely expressed in terms of the part $\boldsymbol{\mathcal{P}}\boldsymbol{x}_d\left(t\right)$ prescribed by the experimenter.

Using the complementary projectors $\boldsymbol{\mathcal{P}}$ and $\boldsymbol{\mathcal{Q}}$, the state matrix $\boldsymbol{\mathcal{A}}$ can be split up in four parts as

$$\boldsymbol{\mathcal{A}} = \boldsymbol{\mathcal{P}}\boldsymbol{\mathcal{A}}\boldsymbol{\mathcal{P}} + \boldsymbol{\mathcal{P}}\boldsymbol{\mathcal{A}}\boldsymbol{\mathcal{Q}} + \boldsymbol{\mathcal{Q}}\boldsymbol{\mathcal{A}}\boldsymbol{\mathcal{P}} + \boldsymbol{\mathcal{Q}}\boldsymbol{\mathcal{A}}\boldsymbol{\mathcal{Q}}. \tag{2.157}$$

Note that the controllability matrix $\tilde{\boldsymbol{\mathcal{K}}}$, Eq. (2.139), does only depend on the parts $\boldsymbol{\mathcal{Q}}\boldsymbol{\mathcal{A}}\boldsymbol{\mathcal{P}}$ and $\boldsymbol{\mathcal{Q}}\boldsymbol{\mathcal{A}}\boldsymbol{\mathcal{Q}}$, but not on $\boldsymbol{\mathcal{P}}\boldsymbol{\mathcal{A}}\boldsymbol{\mathcal{P}}$ and $\boldsymbol{\mathcal{P}}\boldsymbol{\mathcal{A}}\boldsymbol{\mathcal{Q}}$. This fact extends the validity of the controllability matrix $\tilde{\boldsymbol{\mathcal{K}}}$ to nonlinear systems satisfying the linearizing assumption. Furthermore, only the parts $\boldsymbol{\mathcal{Q}}\boldsymbol{\mathcal{A}}\boldsymbol{\mathcal{P}}$ and $\boldsymbol{\mathcal{Q}}\boldsymbol{\mathcal{A}}\boldsymbol{\mathcal{Q}}$ must be known to decide if a system is controllable. Thus, it can be possible to decide about controllability of a system without knowing all details of its dynamics. This insight might be useful for experimental systems with incomplete or approximated model equations.

2.5. Output controllability

Consider the dynamical system

$$\dot{\boldsymbol{x}}(t) = \boldsymbol{R}(\boldsymbol{x}(t)) + \boldsymbol{\mathcal{B}}(\boldsymbol{x}(t))\,\boldsymbol{u}(t), \tag{2.158}$$

together with the output

$$\boldsymbol{z}(t) = \boldsymbol{h}(\boldsymbol{x}(t)). \tag{2.159}$$

Here, $\boldsymbol{z}(t) = (z_1(t), \ldots, z_m(t))^T \in \mathbb{R}^m$ with $m \leq n$ components is called the *output vector* and the *output function* $\boldsymbol{h}$ maps from $\mathbb{R}^n$ to $\mathbb{R}^m$.

A system is called output controllable if it is possible to achieve a transfer from an initial output state

$$\boldsymbol{z}(t_0) = \boldsymbol{z}_0 \tag{2.160}$$

at time $t = t_0$ to a terminal output state

$$\boldsymbol{z}(t_1) = \boldsymbol{z}_1 \tag{2.161}$$

at the terminal time $t = t_1$. In contrast to output controllability, the notion of controllability discussed in Section 2.4 is concerned with the controllability of the state $\boldsymbol{x}(t)$ and is often referred to as state or full state controllability. Note that a state controllable system is not necessarily output controllable. Similarly, an output controllable system is not necessarily state controllable. For $m = n$ and an output function equal to the identity function, $\boldsymbol{h}(\boldsymbol{x}) = \boldsymbol{x}$, output controllability is equivalent to state controllability.

2.5.1. Kalman rank condition for the output controllability of LTI systems

The notion of state controllability developed in form of a Kalman rank condition for an output controllability matrix can be adapted to output controllability (Kalman, 1959, 1960; Chen, 1998). Consider the LTI system with n-dimensional state vector $\boldsymbol{x}(t)$ and p-dimensional control signal $\boldsymbol{u}(t)$,

$$\dot{\boldsymbol{x}}(t) = \boldsymbol{\mathcal{A}}\boldsymbol{x}(t) + \boldsymbol{\mathcal{B}}\boldsymbol{u}(t). \tag{2.162}$$

The output is assumed to be a linear relation of the form

$$\boldsymbol{z}(t) = \boldsymbol{\mathcal{C}}\boldsymbol{x}(t) \tag{2.163}$$

with $m \times n$ *output matrix* $\boldsymbol{\mathcal{C}}$. The $m \times np$ *output controllability matrix* is defined as

$$\boldsymbol{\mathcal{K}}_{\boldsymbol{\mathcal{C}}} = \left(\boldsymbol{\mathcal{C}}\boldsymbol{\mathcal{B}}|\boldsymbol{\mathcal{C}}\boldsymbol{\mathcal{A}}\boldsymbol{\mathcal{B}}|\boldsymbol{\mathcal{C}}\boldsymbol{\mathcal{A}}^2\boldsymbol{\mathcal{B}}|\cdots|\boldsymbol{\mathcal{C}}\boldsymbol{\mathcal{A}}^{n-1}\boldsymbol{\mathcal{B}}\right). \tag{2.164}$$

The LTI system Eq. (2.162) with output (2.163) is output controllable if $\boldsymbol{\mathcal{K}}_{\boldsymbol{\mathcal{C}}}$ satisfies the rank condition

$$\operatorname{rank}\left(\boldsymbol{\mathcal{K}}_{\boldsymbol{\mathcal{C}}}\right) = m. \tag{2.165}$$

A proof of Eq. (2.165) proceeds along the same lines as the proof for the Kalman rank condition for state controllability in Section 2.4.2. Using Eq. (2.94), the solution for $\boldsymbol{z}(t)$ is

$$\boldsymbol{z}(t) = \boldsymbol{\mathcal{C}}\boldsymbol{x}(t) = \boldsymbol{\mathcal{C}}e^{\boldsymbol{\mathcal{A}}(t-t_0)}\boldsymbol{x}_0 + \boldsymbol{\mathcal{C}}\int_{t_0}^{t} d\tau e^{\boldsymbol{\mathcal{A}}(t-\tau)}\boldsymbol{\mathcal{B}}\boldsymbol{u}(\tau). \tag{2.166}$$

Evaluating Eq. (2.166) at the terminal time $t = t_1$ and enforcing the terminal output condition Eq. (2.161) yields a condition for the control signal $\boldsymbol{u}$,

$$\boldsymbol{z}_1 = \boldsymbol{z}(t_1) = \boldsymbol{\mathcal{C}}e^{\boldsymbol{\mathcal{A}}(t_1-t_0)}\boldsymbol{x}_0 + \boldsymbol{\mathcal{C}}\int_{t_0}^{t_1} d\tau e^{\boldsymbol{\mathcal{A}}(t_1-\tau)}\boldsymbol{\mathcal{B}}\boldsymbol{u}(\tau). \tag{2.167}$$

Exploiting the Cayley-Hamilton theorem and proceeding analogously to Eq. (2.97) yields

$$\boldsymbol{z}_1 - \boldsymbol{\mathcal{C}}e^{\boldsymbol{\mathcal{A}}(t_1-t_0)}\boldsymbol{x}_0 = \boldsymbol{\mathcal{C}}\sum_{k=0}^{n-1}\boldsymbol{\mathcal{A}}^k\boldsymbol{\mathcal{B}}\tilde{\boldsymbol{\beta}}_k(t_1,t_0) = \boldsymbol{\mathcal{K}}_{\boldsymbol{\mathcal{C}}}\tilde{\boldsymbol{\beta}}(t_1,t_0) \tag{2.168}$$

with $p \times 1$ vectors $\tilde{\boldsymbol{\beta}}_k$ defined as

$$\tilde{\boldsymbol{\beta}}_k(t_1,t_0) = \int_{t_0}^{t_1} d\tau \frac{(t_1-\tau)^k}{k!}\boldsymbol{u}(\tau) + \sum_{i=n}^{\infty} c_{ik}\int_{t_0}^{t_1} d\tau \frac{(t_1-\tau)^i}{i!}\boldsymbol{u}(\tau) \tag{2.169}$$

and the $np \times 1$ vector

$$\tilde{\boldsymbol{\beta}}(t_1, t_0) = \begin{pmatrix} \tilde{\boldsymbol{\beta}}_0(t_1, t_0) \\ \vdots \\ \tilde{\boldsymbol{\beta}}_{n-1}(t_1, t_0) \end{pmatrix}. \tag{2.170}$$

The linear map from $\tilde{\boldsymbol{\beta}}(t_1, t_0)$ to $\boldsymbol{z}_1$ must be *surjective*. This is the case if and only if the matrix $\boldsymbol{\mathcal{K}}_{\mathcal{C}}$ defined in Eq. (2.164) has full row rank, i.e.,

$$\operatorname{rank}(\boldsymbol{\mathcal{K}}_{\mathcal{C}}) = m. \tag{2.171}$$

2.5.2. Output controllability for systems satisfying the linearizing assumption

Using the framework of exactly realizable trajectories, we can generalize the condition for output controllability in form of a matrix rank condition to nonlinear affine control systems satisfying the linearizing assumption from Section 2.3. Consider the affine control system

$$\dot{\boldsymbol{x}}(t) = \boldsymbol{R}(\boldsymbol{x}(t)) + \boldsymbol{\mathcal{B}}(\boldsymbol{x}(t))\,\boldsymbol{u}(t) \tag{2.172}$$

with linear output

$$\boldsymbol{z}(t) = \boldsymbol{\mathcal{C}}\boldsymbol{x}(t). \tag{2.173}$$

The constraint equation for exactly realizable desired trajectories $\boldsymbol{x}_d(t)$ is linear,

$$\boldsymbol{\mathcal{Q}}\dot{\boldsymbol{x}}_d(t) = \boldsymbol{\mathcal{Q}}\boldsymbol{\mathcal{A}}\boldsymbol{x}_d(t) + \boldsymbol{\mathcal{Q}}\boldsymbol{b}, \tag{2.174}$$

and has the solution

$$\begin{aligned} \boldsymbol{\mathcal{Q}}\boldsymbol{x}_d(t) = {} & \exp(\boldsymbol{\mathcal{Q}}\boldsymbol{\mathcal{A}}\boldsymbol{\mathcal{Q}}(t - t_0))\,\boldsymbol{\mathcal{Q}}\boldsymbol{x}_0 \\ & + \int_{t_0}^{t} d\tau \exp(\boldsymbol{\mathcal{Q}}\boldsymbol{\mathcal{A}}\boldsymbol{\mathcal{Q}}(t - \tau))\,\boldsymbol{\mathcal{Q}}(\boldsymbol{\mathcal{A}}\boldsymbol{\mathcal{P}}\boldsymbol{x}_d(\tau) + \boldsymbol{b}). \end{aligned} \tag{2.175}$$

Enforcing the desired output value at the terminal time $t = t_1$ yields

$$\begin{aligned} \boldsymbol{z}_1 = \boldsymbol{z}_d(t_1) &= \boldsymbol{\mathcal{C}}\boldsymbol{x}_d(t_1) = \boldsymbol{\mathcal{C}}\boldsymbol{\mathcal{P}}\boldsymbol{x}_d(t_1) + \boldsymbol{\mathcal{C}}\boldsymbol{\mathcal{Q}}\boldsymbol{x}_d(t_1) \\ &= \boldsymbol{\mathcal{C}}\boldsymbol{\mathcal{P}}\boldsymbol{x}_d(t_1) + \boldsymbol{\mathcal{C}}\boldsymbol{\mathcal{Q}}\exp(\boldsymbol{\mathcal{Q}}\boldsymbol{\mathcal{A}}\boldsymbol{\mathcal{Q}}(t_1 - t_0))\,\boldsymbol{\mathcal{Q}}\boldsymbol{x}_0 \\ &\quad + \boldsymbol{\mathcal{C}}\boldsymbol{\mathcal{Q}}\int_{t_0}^{t_1} d\tau \exp(\boldsymbol{\mathcal{Q}}\boldsymbol{\mathcal{A}}\boldsymbol{\mathcal{Q}}(t_1 - \tau))\,\boldsymbol{\mathcal{Q}}(\boldsymbol{\mathcal{A}}\boldsymbol{\mathcal{P}}\boldsymbol{x}_d(\tau) + \boldsymbol{b}). \end{aligned} \tag{2.176}$$

That is a condition for the part $\boldsymbol{\mathcal{P}}\boldsymbol{x}_d(\tau)$. Exploiting the Cayley-Hamilton theorem and proceeding as in Eq. (2.134) yields

$$\begin{aligned}\boldsymbol{z}_1 - \boldsymbol{\mathcal{C}}\boldsymbol{\mathcal{Q}}\exp\left(\boldsymbol{\mathcal{Q}}\boldsymbol{\mathcal{A}}\boldsymbol{\mathcal{Q}}\left(t_1-t_0\right)\right)\boldsymbol{\mathcal{Q}}\boldsymbol{x}_0 - \boldsymbol{\mathcal{C}}\boldsymbol{\mathcal{Q}}\int_{t_0}^{t_1} d\tau \exp\left(\boldsymbol{\mathcal{Q}}\boldsymbol{\mathcal{A}}\boldsymbol{\mathcal{Q}}\left(t_1-\tau\right)\right)\boldsymbol{\mathcal{Q}}\boldsymbol{b} \\ = \boldsymbol{\mathcal{C}}\boldsymbol{\mathcal{P}}\boldsymbol{x}_d(t_1) + \boldsymbol{\mathcal{C}}\boldsymbol{\mathcal{Q}}\sum_{k=0}^{n-1}\left(\boldsymbol{\mathcal{Q}}\boldsymbol{\mathcal{A}}\boldsymbol{\mathcal{Q}}\right)^k\tilde{\boldsymbol{\alpha}}_k(t_1,t_0)\,. \quad (2.177)\end{aligned}$$

In Eq. (2.177) we defined the $n\times 1$ vectors

$$\tilde{\boldsymbol{\alpha}}_k(t_1,t_0) = \int_{t_0}^{t_1} d\tau\left(\frac{(t_1-\tau)^k}{k!} + \sum_{i=n}^{\infty} d_{ik}\frac{(t_1-\tau)^i}{i!}\right)\boldsymbol{\mathcal{Q}}\boldsymbol{\mathcal{A}}\boldsymbol{\mathcal{P}}\boldsymbol{x}_d(\tau)\,. \quad (2.178)$$

The right hand side of Eq. (2.177) can be written with the help of the $n(n+1)\times 1$ vector

$$\tilde{\boldsymbol{\alpha}}(t_1,t_0) = \begin{pmatrix}\boldsymbol{\mathcal{P}}\boldsymbol{x}_d(t_1)\\ \boldsymbol{\alpha}_0(t_1,t_0)\\ \vdots\\ \boldsymbol{\alpha}_{n-1}(t_1,t_0)\end{pmatrix} \quad (2.179)$$

and the $m\times n(n+1)$ *output controllability matrix*

$$\tilde{\boldsymbol{\mathcal{K}}}_{\boldsymbol{\mathcal{C}}} = \left(\boldsymbol{\mathcal{C}}\boldsymbol{\mathcal{P}}|\boldsymbol{\mathcal{C}}\boldsymbol{\mathcal{Q}}\boldsymbol{\mathcal{A}}\boldsymbol{\mathcal{P}}|\cdots|\boldsymbol{\mathcal{C}}\boldsymbol{\mathcal{Q}}\left(\boldsymbol{\mathcal{Q}}\boldsymbol{\mathcal{A}}\boldsymbol{\mathcal{Q}}\right)^{n-1}\boldsymbol{\mathcal{Q}}\boldsymbol{\mathcal{A}}\boldsymbol{\mathcal{P}}\right) \quad (2.180)$$

as

$$\begin{aligned}\boldsymbol{z}_1 - \boldsymbol{\mathcal{C}}\boldsymbol{\mathcal{Q}}\exp\left(\boldsymbol{\mathcal{Q}}\boldsymbol{\mathcal{A}}\boldsymbol{\mathcal{Q}}\left(t_1-t_0\right)\right)\boldsymbol{\mathcal{Q}}\boldsymbol{x}_0 - \boldsymbol{\mathcal{C}}\boldsymbol{\mathcal{Q}}\int_{t_0}^{t_1} d\tau \exp\left(\boldsymbol{\mathcal{Q}}\boldsymbol{\mathcal{A}}\boldsymbol{\mathcal{Q}}\left(t_1-\tau\right)\right)\boldsymbol{\mathcal{Q}}\boldsymbol{b} \\ = \tilde{\boldsymbol{\mathcal{K}}}_{\boldsymbol{\mathcal{C}}}\tilde{\boldsymbol{\alpha}}(t_1,t_0)\,. \quad (2.181)\end{aligned}$$

The linear map from $\tilde{\boldsymbol{\alpha}}(t_1,t_0)$ to $\boldsymbol{z}_1$ is surjective if $\tilde{\boldsymbol{\mathcal{K}}}_{\boldsymbol{\mathcal{C}}}$ has full row rank, i.e., if

$$\operatorname{rank}\left(\tilde{\boldsymbol{\mathcal{K}}}_{\boldsymbol{\mathcal{C}}}\right) = m. \quad (2.182)$$

Thus, a nonlinear affine control system satisfying the linearizing assumption is output controllable with linear output Eq. (2.173) if the matrix $\boldsymbol{\mathcal{K}}_{\boldsymbol{\mathcal{C}}}$ satisfies the *output controllability rank condition* Eq. (2.182).

With $m=n$ and $\boldsymbol{\mathcal{C}}=\mathbf{1}$, the notion of output controllability reduces to the notion of full state controllability. Indeed, note that for $\boldsymbol{\mathcal{C}}=\mathbf{1}$, $\tilde{\boldsymbol{\mathcal{K}}}_{\boldsymbol{\mathcal{C}}}$ can be written in terms of the controllability matrix for realizable trajectories $\tilde{\boldsymbol{\mathcal{K}}}$ given by Eq. (2.139) as

$$\tilde{\boldsymbol{\mathcal{K}}}_{\boldsymbol{\mathcal{C}}} = \left(\boldsymbol{\mathcal{P}}|\tilde{\boldsymbol{\mathcal{K}}}\right). \quad (2.183)$$

Because $\tilde{\boldsymbol{\mathcal{K}}}$ has no components in the direction of $\boldsymbol{\mathcal{P}}$, i.e., $\tilde{\boldsymbol{\mathcal{K}}} = \boldsymbol{\mathcal{P}}\tilde{\boldsymbol{\mathcal{K}}} + \boldsymbol{\mathcal{Q}}\tilde{\boldsymbol{\mathcal{K}}} = \boldsymbol{\mathcal{Q}}\tilde{\boldsymbol{\mathcal{K}}}$, the matrix $\tilde{\boldsymbol{\mathcal{K}}}_{\boldsymbol{\mathcal{C}}}$ as given by Eq. (2.183) has rank

$$\operatorname{rank}\left(\tilde{\boldsymbol{\mathcal{K}}}_{\boldsymbol{\mathcal{C}}}\right) = p + \operatorname{rank}\left(\tilde{\boldsymbol{\mathcal{K}}}\right) = n. \tag{2.184}$$

This proves that the rank condition for output controllability, Eq. (2.182), indeed reduces, for $\boldsymbol{\mathcal{C}} = \mathbf{1}$, to the rank condition for full state controllability as given by Eq. (2.140).

Output controllability is discussed by means of two examples.

Example 2.13: Output controllability of the activator-controlled FHN model

The model

$$\begin{pmatrix} \dot{x}(t) \\ \dot{y}(t) \end{pmatrix} = \begin{pmatrix} a_0 + a_1 x(t) + a_2 y(t) \\ R(x(t), y(t)) \end{pmatrix} + \begin{pmatrix} 0 \\ 1 \end{pmatrix} u(t) \tag{2.185}$$

satisfies the linearizing assumption such that the constraint equation is linear with state matrix

$$\boldsymbol{\mathcal{Q}}\boldsymbol{\mathcal{A}} = \begin{pmatrix} a_1 & a_2 \\ 0 & 0 \end{pmatrix}, \tag{2.186}$$

see Examples 2.5 and 2.11 for more details. We check for the controllability of a general desired output with 1×2 output matrix $\boldsymbol{\mathcal{C}} = \begin{pmatrix} c_1, & c_2, \end{pmatrix}^T$,

$$z_d(t) = \boldsymbol{\mathcal{C}}\boldsymbol{x}_d(t) = c_1 x_d(t) + c_2 y_d(t). \tag{2.187}$$

The 1×6 output controllability matrix $\tilde{\boldsymbol{\mathcal{K}}}_{\boldsymbol{\mathcal{C}}}$ becomes

$$\begin{aligned} \tilde{\boldsymbol{\mathcal{K}}}_{\boldsymbol{\mathcal{C}}} &= (\boldsymbol{\mathcal{C}}\boldsymbol{\mathcal{P}} | \boldsymbol{\mathcal{C}}\boldsymbol{\mathcal{Q}}\boldsymbol{\mathcal{A}}\boldsymbol{\mathcal{P}} | \boldsymbol{\mathcal{C}}\boldsymbol{\mathcal{Q}}\boldsymbol{\mathcal{A}}\boldsymbol{\mathcal{Q}}\boldsymbol{\mathcal{A}}\boldsymbol{\mathcal{P}}) \\ &= \begin{pmatrix} 0 & c_2 & 0 & a_2 c_1 & 0 & a_1 a_2 c_1 \end{pmatrix}. \end{aligned} \tag{2.188}$$

The rank of $\tilde{\boldsymbol{\mathcal{K}}}_{\boldsymbol{\mathcal{C}}}$ is at most one. Example 2.11 showed that the system Eq. (2.185) is not controllable if $a_2 = 0$. In this case, $\tilde{\boldsymbol{\mathcal{K}}}_{\boldsymbol{\mathcal{C}}}$ simplifies to

$$\tilde{\boldsymbol{\mathcal{K}}}_{\boldsymbol{\mathcal{C}}} = \begin{pmatrix} 0 & c_2 & 0 & 0 & 0 & 0 \end{pmatrix}. \tag{2.189}$$

Thus, $\tilde{\boldsymbol{\mathcal{K}}}_{\boldsymbol{\mathcal{C}}}$ still has rank one as long as $c_2 \neq 0$. In conclusion, a model which is not controllable can nevertheless have a controllable output. Although for $a_2 = 0$ in Eq. (2.185), the inhibitor $x(t)$ evolves uncoupled from the activator

dynamics, activator and inhibitor are still coupled in the output $z_d(t)$, Eq. (2.187). In that way the activator $y_d(t)$ can counteract the inhibitor $x_d(t)$ to control the desired output. If additionally $c_2 = 0$, this is not possible, and the output is not controllable. Indeed, for $c_2 = 0$, the output controllability matrix Eq. (2.189) reduces to the zero matrix with vanishing rank.

Example 2.14: Output controllability of the SIR model

The controlled state equation for the SIR model was developed in Example 1.3 and is repeated here for convenience,

$$\dot{S}(t) = -(\beta + u(t))\frac{S(t)I(t)}{N}, \qquad \dot{I}(t) = (\beta + u(t))\frac{S(t)I(t)}{N} - \gamma I(t),$$
$$\dot{R}(t) = \gamma I(t). \tag{2.190}$$

The controllability of the SIR model was discussed in Example 2.12. We check for the controllability of a general single component desired output with 1×3 output matrix $\boldsymbol{\mathcal{C}} = \begin{pmatrix} c_1, & c_2, & c_3 \end{pmatrix}^T$,

$$z_d(t) = \boldsymbol{\mathcal{C}}\boldsymbol{x}_d(t) = c_1 S_d(t) + c_2 I_d(t) + c_3 R_d(t). \tag{2.191}$$

The output controllability matrix $\tilde{\boldsymbol{\mathcal{K}}}_{\boldsymbol{\mathcal{C}}}$ becomes

$$\begin{aligned}\tilde{\boldsymbol{\mathcal{K}}}_{\boldsymbol{\mathcal{C}}} &= (\boldsymbol{\mathcal{C}}\boldsymbol{\mathcal{P}}|\boldsymbol{\mathcal{C}}\boldsymbol{\mathcal{Q}}\boldsymbol{\mathcal{A}}\boldsymbol{\mathcal{P}}|\boldsymbol{\mathcal{C}}\boldsymbol{\mathcal{Q}}\boldsymbol{\mathcal{A}}\boldsymbol{\mathcal{Q}}\boldsymbol{\mathcal{A}}\boldsymbol{\mathcal{P}}|\boldsymbol{\mathcal{C}}\boldsymbol{\mathcal{Q}}\boldsymbol{\mathcal{A}}\boldsymbol{\mathcal{Q}}\boldsymbol{\mathcal{A}}\boldsymbol{\mathcal{Q}}\boldsymbol{\mathcal{A}}\boldsymbol{\mathcal{P}}) \\ &= \begin{pmatrix} \kappa_1 & -\kappa_1 & 0 & \gamma\kappa_2 & -\gamma\kappa_2 & 0 & -\frac{1}{2}\gamma^2\kappa_2 & \frac{\gamma^2}{2}\kappa_2 & 0 & \frac{\gamma^3}{4}\kappa_2 & -\frac{1}{4}\gamma^3\kappa_2 & 0 \end{pmatrix},\end{aligned} \tag{2.192}$$

with $\kappa_1 = \frac{1}{2}(c_1 - c_2)$ and $\kappa_2 = \frac{1}{4}(c_1 + c_2 - 2c_3)$. The rank of $\tilde{\boldsymbol{\mathcal{K}}}_{\boldsymbol{\mathcal{C}}}$ is at most one. Depending on the values of the output parameters c_1, c_2, c_3, and the system parameter γ, the rank of $\tilde{\boldsymbol{\mathcal{K}}}_{\boldsymbol{\mathcal{C}}}$ changes. Two cases are discussed in detail.

First, if $c_1 = c_2 = 0$ and $c_3 \neq 0$, then $\kappa_1 = 0$ and the output is $z_d(t) = c_3 R_d(t)$ and prescribes the number of recovered people over time. As can be seen from Eq. (2.190), $R(t)$ is decoupled from the controlled part of the equations if $\gamma = 0$. Indeed, in this case $\tilde{\boldsymbol{\mathcal{K}}}_{\boldsymbol{\mathcal{N}}}$ reduces to the zero matrix with vanishing rank. In conclusion, a desired output equal to the number $R_d(t)$ of recovered people cannot be controlled if $\gamma = 0$.

Second, for $c_1 = c_2 = c_3 = c$ the desired output becomes

$$z_d(t) = c(S_d(t) + I_d(t) + R_d(t)) = cN = \text{const.}, \tag{2.193}$$

with N being the total number of individuals. This conservation law can easily be derived from the system dynamics Eq. (2.190) and remains true for the controlled system. We expect that this output is not controllable because the value of N is fixed by the initial conditions and cannot be changed by control. Indeed, if $c_1 = c_2 = c_3$ then $\kappa_1 = 0$ and $\kappa_2 = 0$. The output controllability matrix $\tilde{\boldsymbol{\mathcal{K}}}_{\boldsymbol{\mathcal{C}}}$ becomes the zero matrix with vanishing rank, and the output Eq. (2.193) is not controllable.

2.6. Output realizability

2.6.1. General procedure

For a desired trajectory $\boldsymbol{x}_d(t)$ to be exactly realizable, it must satisfy the constraint equation

$$\boldsymbol{\mathcal{Q}}(\boldsymbol{x}_d(t))\left(\dot{\boldsymbol{x}}_d(t) - \boldsymbol{R}(\boldsymbol{x}_d(t))\right) = \boldsymbol{0}. \tag{2.194}$$

This equation fixes $n-p$ components of the n components of $\boldsymbol{x}_d(t)$. Our convention was to choose these $n-p$ independent components as $\boldsymbol{y}_d(t) = \boldsymbol{\mathcal{Q}}(\boldsymbol{x}_d(t))\,\boldsymbol{x}_d(t)$, while the p independent components $\boldsymbol{z}_d(t) = \boldsymbol{\mathcal{P}}(\boldsymbol{x}_d(t))\,\boldsymbol{x}_d(t)$ of the desired state trajectory are prescribed by the experimenter. Equation (2.194) becomes a non-autonomous differential equation for $\boldsymbol{y}_d(t)$,

$$\begin{aligned}\dot{\boldsymbol{y}}_d(t) = {} & \boldsymbol{\mathcal{Q}}(\boldsymbol{y}_d(t) + \boldsymbol{z}_d(t))\,\boldsymbol{R}(\boldsymbol{y}_d(t) + \boldsymbol{z}_d(t)) \\ & + \dot{\boldsymbol{\mathcal{Q}}}(\boldsymbol{y}_d(t) + \boldsymbol{z}_d(t))\,(\boldsymbol{y}_d(t) + \boldsymbol{z}_d(t)).\end{aligned} \tag{2.195}$$

Here, $\dot{\boldsymbol{\mathcal{Q}}}$ denotes the short hand notation

$$\dot{\boldsymbol{\mathcal{Q}}}(\boldsymbol{x}(t)) = \frac{d}{dt}\boldsymbol{\mathcal{Q}}(\boldsymbol{x}(t)) = \left(\dot{\boldsymbol{x}}^T(t)\,\nabla\right)\boldsymbol{\mathcal{Q}}(\boldsymbol{x}(t)). \tag{2.196}$$

The explicit time dependence rendering Eq. (2.195) a non-autonomous differential equation comes from the term $\boldsymbol{z}_d(t)$. The initial condition for Eq. (2.195) is

$$\boldsymbol{y}_d(t_0) = \boldsymbol{\mathcal{Q}}(\boldsymbol{x}(t_0))\,\boldsymbol{x}(t_0), \tag{2.197}$$

while $\boldsymbol{z}_d(t_0)$ has to satisfy

$$\boldsymbol{z}_d(t_0) = \boldsymbol{\mathcal{P}}(\boldsymbol{x}(t_0))\,\boldsymbol{x}(t_0). \tag{2.198}$$

Because of $\boldsymbol{\mathcal{B}}^+(\boldsymbol{x}_d(t))\,\boldsymbol{\mathcal{Q}}(\boldsymbol{x}_d(t)) = \boldsymbol{0}$, the corresponding control signal $\boldsymbol{u}(t)$ is given as

$$\begin{aligned}\boldsymbol{u}(t) &= \boldsymbol{\mathcal{B}}^+(\boldsymbol{x}_d(t))\left(\dot{\boldsymbol{x}}_d(t) - \boldsymbol{R}(\boldsymbol{x}_d(t))\right) \\ &= \boldsymbol{\mathcal{B}}^+(\boldsymbol{y}_d(t) + \boldsymbol{z}_d(t))\left(\dot{\boldsymbol{z}}_d(t) - \boldsymbol{R}(\boldsymbol{y}_d(t) + \boldsymbol{z}_d(t))\right) \\ &\quad - \boldsymbol{\mathcal{B}}^+(\boldsymbol{y}_d(t) + \boldsymbol{z}_d(t))\,\dot{\boldsymbol{\mathcal{P}}}(\boldsymbol{y}_d(t) + \boldsymbol{z}_d(t))\,(\boldsymbol{y}_d(t) + \boldsymbol{z}_d(t)).\end{aligned} \tag{2.199}$$

Solving Eq. (2.195) for $\boldsymbol{y}_d(t)$ in terms of $\boldsymbol{z}_d(t)$, the term $\boldsymbol{y}_d(t)$ is eliminated from Eq. (2.199), resulting in a control signal expressed in terms of $\boldsymbol{z}_d(t)$ only. The dependence of the control signal $\boldsymbol{u}(t)$ on $\boldsymbol{z}_d(t)$ is in form of a functional,

$$\boldsymbol{u}(t) = \boldsymbol{u}[\boldsymbol{z}_d(t)]. \tag{2.200}$$

However, the choice $\boldsymbol{z}_d(t) = \boldsymbol{\mathcal{P}}(\boldsymbol{x}_d(t))\,\boldsymbol{x}_d(t)$ is not the only possible desired output. A general approach prescribes an arbitrary m-component output

$$\boldsymbol{z}_d(t) = \boldsymbol{h}(\boldsymbol{x}_d(t)). \tag{2.201}$$

The function $\boldsymbol{h}$ maps the state space $\mathbb{R}^n$ to a space $\mathbb{R}^m$. Using the constraint equation (2.194), one can attempt to eliminate $n-m$ components of $\boldsymbol{x}_d(t)$ in the control signal and obtain a control signal depending on $\boldsymbol{z}_d(t)$ only. If it is possible to do so, also the controlled state trajectory $\boldsymbol{x}(t)$ can be expressed in terms of the desired output $\boldsymbol{z}_d(t)$ only. The output $\boldsymbol{z}_d(t)$ is an *exactly realizable desired output*. Clearly, not all desired outputs can be realized, and the question arises under which conditions it is possible to exactly realize a desired output $\boldsymbol{z}_d(t)$. For example, if the dimension m of the output signals is larger than the dimension p of the control signals, $m > p$, it should be impossible to express the control signal in terms of the output. Here, we are not able to give a definite answer to this question. We discuss some general aspects of the problem in Section 2.6.2, and treat some explicit examples in Sections 2.6.3.

A remark in order to minimize the confusion: $\boldsymbol{z}_d$ as given by $\boldsymbol{z}_d = \boldsymbol{\mathcal{P}}(\boldsymbol{x}_d)\,\boldsymbol{x}_d$ is an n-component vector, but has only p independent components. Starting with Eq. (2.201), the output $\boldsymbol{z}_d(t)$ is regarded as a p-component vector with p independent components, as it is customary for outputs. Note that the convention of choosing the part $\boldsymbol{\mathcal{P}}\boldsymbol{x}_d(t)$ as the desired output corresponds to the case $\boldsymbol{\mathcal{M}} = \boldsymbol{\mathcal{P}}$ and $\boldsymbol{\mathcal{N}} = \boldsymbol{\mathcal{Q}}$, which, for constant coupling matrix $\boldsymbol{\mathcal{B}}(\boldsymbol{x}) = \boldsymbol{\mathcal{B}}$, is equivalent to the linear output function $\boldsymbol{z}(t) = \boldsymbol{\mathcal{B}}^T\boldsymbol{x}(t)$.

2.6.2. Output trajectory realizability leads to differential-algebraic systems

Consider a desired output trajectory $\boldsymbol{z}_d(t)$ depending linearly on the desired state trajectory $\boldsymbol{x}_d(t)$,

$$\boldsymbol{z}_d(t) = \boldsymbol{\mathcal{C}}\boldsymbol{x}_d(t). \tag{2.202}$$

The desired output $\boldsymbol{z}_d(t)$ has $m \leq n$ independent components and $\boldsymbol{\mathcal{C}}$ is assumed to be a constant $m \times n$ *output matrix* with full rank,

$$\operatorname{rank}(\boldsymbol{\mathcal{C}}) = m. \tag{2.203}$$

Equation (2.202) is viewed as an underdetermined system of linear equations for the desired state $\boldsymbol{x}_d(t)$. See the Appendix A.2 for an introduction in solving underdetermined systems of equations.

For the linear output given by Eq. (2.202), we can define two complementary projectors $\boldsymbol{\mathcal{M}}$ and $\boldsymbol{\mathcal{N}}$ by

$$\boldsymbol{\mathcal{M}} = \boldsymbol{\mathcal{C}}^{+}\boldsymbol{\mathcal{C}}, \qquad \boldsymbol{\mathcal{N}} = \mathbf{1} - \boldsymbol{\mathcal{M}}. \tag{2.204}$$

Here, the Moore-Penrose pseudo inverse $\boldsymbol{\mathcal{C}}^{+}$ of $\boldsymbol{\mathcal{C}}$ is given by

$$\boldsymbol{\mathcal{C}}^{+} = \boldsymbol{\mathcal{C}}^{T}\left(\boldsymbol{\mathcal{C}}\boldsymbol{\mathcal{C}}^{T}\right)^{-1}. \tag{2.205}$$

The projectors $\boldsymbol{\mathcal{M}}$ and $\boldsymbol{\mathcal{N}}$ are symmetric $n \times n$ matrices. The inverse of the $m \times m$ matrix $\boldsymbol{\mathcal{C}}\boldsymbol{\mathcal{C}}^{T}$ exists because $\boldsymbol{\mathcal{C}}$ has full rank by assumption. The ranks of the projectors are

$$\operatorname{rank}(\boldsymbol{\mathcal{M}}) = m, \qquad \operatorname{rank}(\boldsymbol{\mathcal{N}}) = n - m. \tag{2.206}$$

Multiplying $\boldsymbol{\mathcal{M}}$ and $\boldsymbol{\mathcal{N}}$ with $\boldsymbol{\mathcal{C}}$ from the left and right yields

$$\boldsymbol{\mathcal{M}}\boldsymbol{\mathcal{C}}^{T} = \boldsymbol{\mathcal{C}}^{T}, \qquad \boldsymbol{\mathcal{C}}\boldsymbol{\mathcal{M}} = \boldsymbol{\mathcal{C}}, \qquad \boldsymbol{\mathcal{N}}\boldsymbol{\mathcal{C}}^{T} = \mathbf{0}, \qquad \boldsymbol{\mathcal{C}}\boldsymbol{\mathcal{N}} = \mathbf{0}. \tag{2.207}$$

Multiplying the state-output relation (2.202) by $\boldsymbol{\mathcal{C}}^{+}$ from the left gives

$$\boldsymbol{\mathcal{M}}\boldsymbol{x}_d(t) = \boldsymbol{\mathcal{C}}^{+}\boldsymbol{z}_d(t). \tag{2.208}$$

Using the last equation, the desired state $\boldsymbol{x}_d(t)$ can be separated in two parts as

$$\boldsymbol{x}_d(t) = \boldsymbol{\mathcal{M}}\boldsymbol{x}_d(t) + \boldsymbol{\mathcal{N}}\boldsymbol{x}_d(t) = \boldsymbol{\mathcal{C}}^{+}\boldsymbol{z}_d(t) + \boldsymbol{\mathcal{N}}\boldsymbol{x}_d(t). \tag{2.209}$$

Thus, the part $\boldsymbol{\mathcal{M}}\boldsymbol{x}_d(t)$ can be expressed in terms of the output $\boldsymbol{z}_d(t)$ while the part $\boldsymbol{\mathcal{N}}\boldsymbol{x}_d(t)$ is left undetermined.

In the following, we enforce the first part of the linearizing assumption, namely, we assume constant projectors

$$\boldsymbol{\mathcal{P}}(\boldsymbol{x}) = \boldsymbol{\mathcal{P}} = \text{const.}, \qquad \boldsymbol{\mathcal{Q}}(\boldsymbol{x}) = \mathbf{1} - \boldsymbol{\mathcal{P}} = \text{const.}, \tag{2.210}$$

in the constraint equation (2.194). The constraint equation becomes

$$\boldsymbol{\mathcal{Q}}\dot{\boldsymbol{x}}_d(t) = \boldsymbol{\mathcal{Q}}\boldsymbol{R}(\boldsymbol{x}_d(t)). \tag{2.211}$$

Using the projectors $\boldsymbol{\mathcal{M}}$ and $\boldsymbol{\mathcal{N}}$ introduced in Eq. (2.209), the constraint equation can be written as the *output constraint equation*

$$\boldsymbol{\mathcal{Q}}\boldsymbol{\mathcal{N}}\dot{\boldsymbol{x}}_d(t) = \boldsymbol{\mathcal{Q}}\boldsymbol{R}\left(\boldsymbol{\mathcal{C}}^{+}\boldsymbol{z}_d(t) + \boldsymbol{\mathcal{N}}\boldsymbol{x}_d(t)\right) - \boldsymbol{\mathcal{Q}}\boldsymbol{\mathcal{C}}^{+}\dot{\boldsymbol{z}}_d(t). \tag{2.212}$$

This is a system of equations for the part $\boldsymbol{\mathcal{N}}\boldsymbol{x}_d(t)$. However, note that the rank of the matrix product $\boldsymbol{\mathcal{Q}}\boldsymbol{\mathcal{N}}$ is

$$\begin{aligned} r &= \operatorname{rank}(\boldsymbol{\mathcal{Q}}\boldsymbol{\mathcal{N}}) \leq \min(\operatorname{rank}(\boldsymbol{\mathcal{Q}}), \operatorname{rank}(\boldsymbol{\mathcal{N}})) \\ &= \min(n-p, n-m). \end{aligned} \tag{2.213}$$

In the most extreme case, $\boldsymbol{\mathcal{Q}}\boldsymbol{\mathcal{N}} = \mathbf{0}$ and so $r = 0$, and Eq. (2.212) reduces to a purely algebraic equation for $\boldsymbol{\mathcal{N}}\boldsymbol{x}_d(t)$,

$$\mathbf{0} = \boldsymbol{\mathcal{Q}}\boldsymbol{R}\left(\boldsymbol{\mathcal{C}}^{+}\boldsymbol{z}_d(t) + \boldsymbol{\mathcal{N}}\boldsymbol{x}_d(t)\right) - \boldsymbol{\mathcal{Q}}\boldsymbol{\mathcal{C}}^{+}\dot{\boldsymbol{z}}_d(t). \tag{2.214}$$

In general, Eq. (2.212) is a system differential-algebraic equations for the part $\boldsymbol{\mathcal{N}}\boldsymbol{x}_d(t)$, and the order of the differential equation depends on the rank of $\boldsymbol{\mathcal{Q}}\boldsymbol{\mathcal{N}}$. For $m = p$, the system consists of r independent differential equations and $n - p - r$ algebraic equations. See the books (Campbell, 1980, 1982; Kunkel and Mehrmann, 2006) for more information about differential-algebraic equations.

Changing the order of differential equations implies consequences for its initial conditions. For example, evaluating Equation (2.214) at the initial time $t = t_0$,

$$\mathbf{0} = \boldsymbol{\mathcal{Q}}\boldsymbol{R}\left(\boldsymbol{\mathcal{C}}^{+}\boldsymbol{z}_d(t_0) + \boldsymbol{\mathcal{N}}\boldsymbol{x}_d(t_0)\right) - \boldsymbol{\mathcal{Q}}\boldsymbol{\mathcal{C}}^{+}\dot{\boldsymbol{z}}_d(t_0), \tag{2.215}$$

uncovers an additional relation between $\boldsymbol{x}_d(t_0)$ and $\boldsymbol{z}_d(t_0)$ which also involves the time derivative $\dot{\boldsymbol{z}}_d(t_0)$. If in an experiment the initial state $\boldsymbol{x}(t_0) = \boldsymbol{x}_0$ of the system can be prepared, Eq. (2.215) yields the value for the part $\boldsymbol{\mathcal{N}}\boldsymbol{x}_d(t_0)$, while the part $\boldsymbol{\mathcal{M}}\boldsymbol{x}_d(t_0)$ is given by

$$\boldsymbol{\mathcal{M}}\boldsymbol{x}_d(t_0) = \boldsymbol{\mathcal{C}}^{+}\boldsymbol{z}_d(t_0). \tag{2.216}$$

On the other hand, if the initial state of the system cannot be prepared, Eq. (2.215) enforces an explicit relation between $\boldsymbol{\mathcal{N}}\boldsymbol{x}_d(t_0)$ and $\dot{\boldsymbol{z}}_d(t_0)$. In general, for $m = p$ and $r = \operatorname{rank}(\boldsymbol{\mathcal{Q}}\boldsymbol{\mathcal{N}}) < n - p$, $n - p - r$ additional conditions have to be satisfied by the initial time derivatives of the desired output trajectory $\boldsymbol{z}_d(t)$. We discuss output realizability with help of several examples.

2.6.3. Realizing a desired output: Examples

Example 2.15: Realizing a desired output for the activator-controlled FHN model

Consider the model

$$\begin{pmatrix} \dot{x}(t) \\ \dot{y}(t) \end{pmatrix} = \begin{pmatrix} a_0 + a_1 x(t) + a_2 y(t) \\ R(x(t), y(t)) \end{pmatrix} + \begin{pmatrix} 0 \\ 1 \end{pmatrix} u(t), \tag{2.217}$$

with nonlinearity

$$R(x,y) = R(y) - x. \tag{2.218}$$

The function $R(y) = y - \frac{1}{3}y^3$ corresponds to the standard FHN nonlinearity. The constraint equation is linear

$$\dot{x}_d(t) = a_0 + a_1 x_d(t) + a_2 y_d(t). \tag{2.219}$$

In Example 2.5, the conventional choice of prescribing the activator variable $y_d(t)$ was applied. The constraint equation (2.219) was regarded as a differential equation for $x_d(t)$. Consequently, by eliminating $x_d(t)$ in the control $u(t) = \dot{y}_d(t) - R(x_d(t), y_d(t))$, $u(t)$ was expressed entirely in terms of $y_d(t)$. In contrast, here the output $z_d(t)$ is chosen as a linear combination of activator and inhibitor,

$$z_d(t) = h(x_d(t), y_d(t)) = c_1 x_d(t) + c_2 y_d(t). \tag{2.220}$$

Rearranging Eq. (2.220) gives

$$y_d(t) = \frac{1}{c_2}(z_d(t) - c_1 x_d(t)). \tag{2.221}$$

Using the last relation in the constraint equation (2.219) yields a linear ODE for x_d with inhomogeneity z_d,

$$\dot{x}_d(t) = \left(a_1 - \frac{c_1 a_2}{c_2}\right) x_d(t) + a_0 + \frac{a_2}{c_2} z_d(t). \tag{2.222}$$

Its solution is, with $\kappa = a_1 - \frac{a_2 c_1}{c_2}$,

$$\begin{aligned} x_d(t) = x_d(t_0)\, e^{\kappa(t-t_0)} + e^{\kappa t}\frac{a_2}{c_2}\int_{t_0}^{t} \exp(-\kappa\tau)\, z_d(\tau)\, d\tau \\ + \frac{a_0}{\kappa}\left(e^{\kappa(t-t_0)} - 1\right), \end{aligned} \tag{2.223}$$

Using relation Eq. (2.221) for y_d together with the Eq. (2.223), x_d and y_d can be eliminated in terms of $z_d(t)$ from the control signal $u(t)$. The result is an expression in terms of the desired output z_d only (not shown).

One remark about the initial condition for the desired state $z_d(t)$. For any desired trajectory $\boldsymbol{x}_d(t)$ to be exactly realizable, its initial condition $\boldsymbol{x}_d(t_0)$ must agree with the initial condition $\boldsymbol{x}(t_0)$ of the controlled state $\boldsymbol{x}(t)$. This naturally restricts the initial value of the desired output to satisfy $z_d(t_0) = c_1 x_d(t_0) + c_2 y_d(t_0)$.

In conclusion, the control as well as the desired state trajectory $\boldsymbol{x}_d(t)$ is expressed solely in terms of the desired output $z_d(t)$. A numerical simulation of the controlled model shown in Fig. 2.5 demonstrates the successful realization of the desired output $z_d(t) = 4\sin(2t)$. Initially at time $t_0 = 0$, the state is set to $x_0 = y_0 = 0$, which complies with the initial value of the desired output $z_d(0) = 0$. A comparison of the desired output $z_d(t)$ with the output $z(t)$ obtained by numerical simulations of the controlled system demonstrates perfect agreement (see Fig. 2.5 top left), and a plot of $z(t) - z_d(t)$ reveals differences within numerical precision (see Fig. 2.5 top right). The controlled state trajectories $x(t)$ and $y(t)$ are shown in Fig. 2.5 bottom left, and the control signal is shown in Fig. 2.5 bottom right.

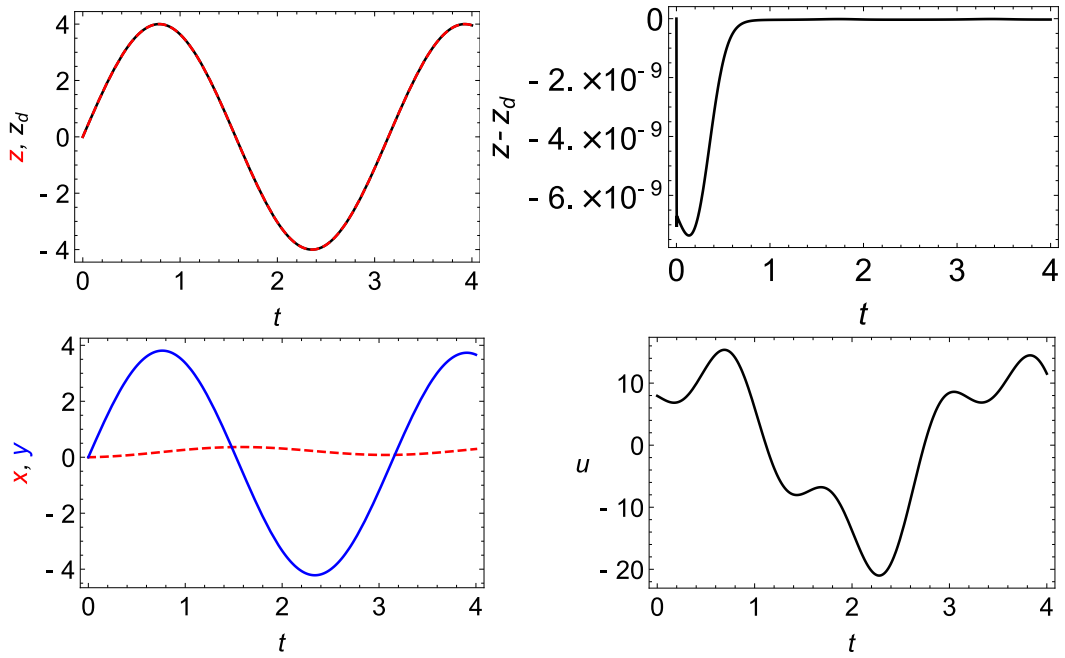

Figure 2.5.: Realizing a desired output in the activator-controlled FHN model. The numerical result z (red dashed line) for the output lies on top of the desired output trajectory z_d (black line), see top left figure. The difference $z - z_d$ is within the range of numerical precision (top right). The bottom left figure shows the corresponding state trajectories x (red dashed line) and y (blue line) and the control u (bottom right).

Example 2.16: Controlling the number of infected individuals in the SIR model

The controlled state equation for the SIR model was developed in Example 1.3, and its output controllability was discussed in Example 2.14. An uncontrolled time evolution is assumed for all times $t < t_0$, upon which the control is switched on. Starting at time $t = t_0$, the number of infected people over time is prescribed. The desired output is

$$z_d(t) = I_d(t). \tag{2.224}$$

The constraint equation consists of two independent equations

$$\begin{pmatrix} \frac{1}{2}\left(\gamma z_d(t) + \dot{z}_d(t) + \dot{S}_d(t)\right) \\ -\gamma z_d(t) + \dot{R}_d(t) \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}. \tag{2.225}$$

The constraint equation is considered as two differential equations for $S_d(t)$ and $R_d(t)$. Their solutions are readily obtained as

$$S_d(t) = -\gamma \int_{t_0}^{t} d\tau z_d(\tau) - z_d(t) + S_d(t_0) + z_d(t_0), \tag{2.226}$$

$$R_d(t) = R_d(t_0) + \gamma \int_{t_0}^{t} d\tau z_d(\tau). \tag{2.227}$$

Eqs. (2.226) and (2.227) express $S_d(t)$ and $R_d(t)$ solely in terms of the desired output $z_d(t)$ and the initial conditions. For any desired trajectory $\boldsymbol{x}_d(t)$ to be exactly realizable, its initial condition $\boldsymbol{x}_d(t_0)$ must comply with the initial condition $\boldsymbol{x}(t_0)$. For the initial conditions of R_d and S_d follows

$$R_d(t_0) = R(t_0), \qquad S_d(t_0) = S(t_0), \tag{2.228}$$

while from $z_d(t) = I_d(t)$ follows

$$z_d(t_0) = I(t_0), \tag{2.229}$$

with $I(t_0)$ being the number of infected people at time $t = t_0$ when control measures are started.

The solution for the control signal realizing a desired trajectory $\boldsymbol{x}_d(t)$ is

$$\begin{aligned} u(t) &= \left(\boldsymbol{B}^T(\boldsymbol{x}_d(t))\,\boldsymbol{B}(\boldsymbol{x}_d(t))\right)^{-1}\boldsymbol{B}^T(\boldsymbol{x}_d(t))\left(\dot{\boldsymbol{x}}_d(t)-\boldsymbol{R}(\boldsymbol{x}_d(t))\right) \\ &= N\frac{\gamma I_d(t)+\dot{I}_d(t)-\dot{S}_d(t)}{2I_d(t)\,S_d(t)}-\beta, \end{aligned} \tag{2.230}$$

and using the solutions for $S_d(t)$ and $R_d(t)$ in terms of $z_d(t)$, the control signal becomes

$$u(t) = N\frac{\gamma z_d(t)+\dot{z}_d(t)}{z_d(t)\left(S(t_0)+I(t_0)-z_d(t)-\gamma\int_{t_0}^{t}d\tau\, z_d(\tau)\right)}-\beta. \tag{2.231}$$

The desired number of infected individuals $z_d(t)$ shall follow a parabolic time evolution,

$$z_d(t) = b_2t^2+b_1t+b_0. \tag{2.232}$$

Three conditions are necessary to determine the three constants b_0, b_1, and b_2. The first condition follows from Eq. (2.229). Second, the number of infected individuals shall vanish at time $t=t_1$,

$$z_d(t_1) = 0, \tag{2.233}$$

such that t_1-t_0 is the duration of the epidemic. To obtain a third relation, we demand that initially, the control signal vanishes. Evaluating Eq. (2.231) at $t=t_0$ yields

$$u(t_0) = N\frac{\gamma I(t_0)+\dot{z}_d(t_0)}{I(t_0)\,S(t_0)}-\beta = 0. \tag{2.234}$$

This relation can be used to obtain a relation for $\dot{z}_d(t_0)$ as

$$\dot{z}_d(t_0) = \frac{\beta}{N}I(t_0)\,S(t_0)-\gamma I(t_0). \tag{2.235}$$

Equation (2.234) guarantees a smooth transition of the time-dependent transmission rate $\beta(t)=\beta+u(t)$ across $t=t_0$.

Figure 2.6 shows a numerical solution. Up to time $t=t_0$, the system evolves uncontrolled, upon which all initial state values $S(t_0)$, $I(t_0)$, and $R(t_0)$ are measured. Starting at time $t_0=10$, the control signal $u(t)$, Eq. (2.231), acts

on the system. To prevent an unphysical negative transmission rate $\beta(t) = \beta + u(t)$, the control $u(t)$ is clipped,

$$\hat{u}(t) = \begin{cases} u(t), & u(t) > -\beta, \\ -\beta, & u(t) \leq -\beta. \end{cases} \tag{2.236}$$

As can be seen in Fig. 2.6 bottom right, $\beta + u(t)$ reaches zero at an approximate time $\tilde{t}_1 \approx 56$, upon which the system evolves again uncontrolled. At this time, the epidemic has reached a reproductive number (see Example 1.3)

$$R_0 = \frac{\beta + u\left(\tilde{t}_1\right)}{\gamma} = 0 < 1, \tag{2.237}$$

and further spreading of the epidemic is prevented. Comparison of the controlled output $z(t) = I(t)$ with its desired counterpart $z_d(t) = I_d(t)$ shows perfect agreement for times $t_0 < t < \tilde{t}_1$ when control measures are operative, see bottom left of Fig. 2.6. Comparing the left and right top figures of Fig. 2.6 reveals a less dramatic epidemic in case of control (top right) than in case without control (top left), with a lower maximum number of infected individuals $I(t)$ (red) and a smaller final number of recovered individuals $R(t)$ (black). Note that $R(t)$ is equivalent to the cumulative number of peoples affected by the epidemic.

While no exact analytical solution is known for the uncontrolled SIR model, we easily managed to find an exact analytical solution for the control as well as for the controlled state over time. This simple analytical approach provides statements as "If the number of infected individuals Δt days from now shall not exceed $I_{\Delta t}$, the transmission rate has to be lowered by $\Delta\beta$ within the next Δt_1 days" without much computational effort. It is a way to predict the effectiveness versus cost of control measures. Of course, application of this result to real world systems requires a model for the cost of quarantine measures or vaccination programs and their impact on the transmission rate $\beta(t)$.

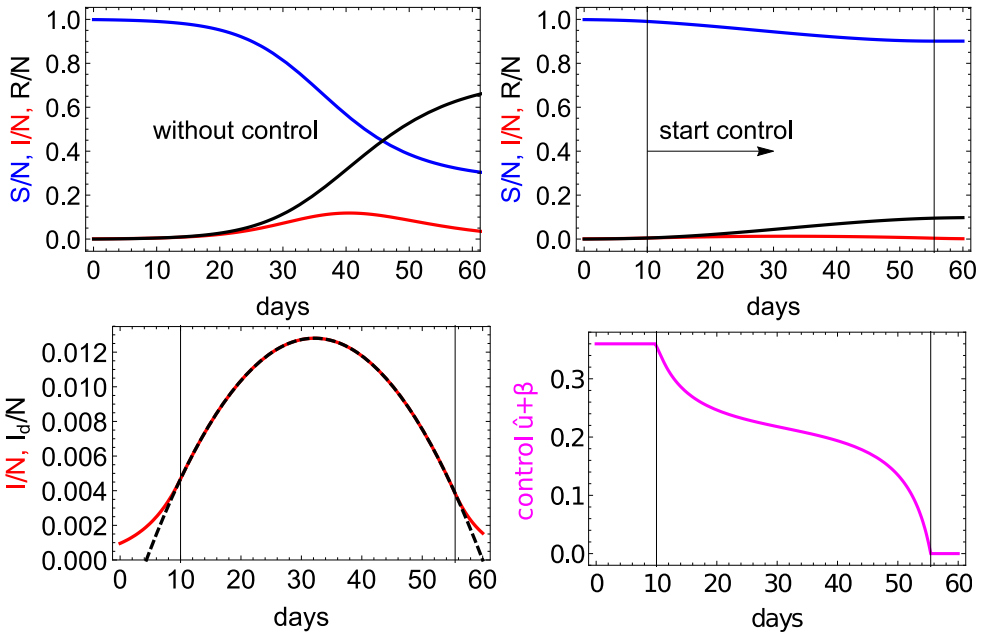

Figure 2.6.: Control of an epidemic in the SIR model. Without control (top left), much more individuals become infected (red) than with control measures starting at $t = 10$ (top right). A comparison of the desired output $z_d(t) = I_d(t)$ (black dashed line) of infected individuals with the actual output trajectory $z(t) = I(t)$ (red solid line) of the controlled dynamical system reveals perfect agreement for times $t_0 < t < \tilde{t}_1$ when control measures are operative (bottom left). The bottom right figure shows the control signal which is clipped such that the time-dependent transmission rate $\beta(t) = \beta + \hat{u}(t) > 0$ is always positive.

Example 2.17: Activator as output for the inhibitor-controlled FHN model

Consider the model with coupling vector $\boldsymbol{B} = \begin{pmatrix} 1, & 0 \end{pmatrix}^T$

$$\begin{pmatrix} \dot{x}(t) \\ \dot{y}(t) \end{pmatrix} = \begin{pmatrix} a_0 + a_1 x(t) + a_2 y(t) \\ R(x(t), y(t)) \end{pmatrix} + \begin{pmatrix} 1 \\ 0 \end{pmatrix} u(t) \tag{2.238}$$

and with nonlinearity

$$R(x, y) = R(y) - x. \tag{2.239}$$

The function $R(y) = y - \frac{1}{3}y^3$ corresponds to the standard FHN nonlinearity. Example 2.6 applied the conventional choice and prescribed the inhibitor variable $x_d(t)$ as the desired output, while $y_d(t)$ was determined as the solution to

the corresponding constraint equation. In contrast, here the desired output is given by the activator $y_d(t)$

$$z_d(t) = y_d(t). \tag{2.240}$$

The control signal in terms of the desired trajectory $\boldsymbol{x}_d(t) = \begin{pmatrix} x_d(t), & y_d(t) \end{pmatrix}^T$ is

$$u(t) = \dot{x}_d(t) - a_0 - a_1 x_d(t) - a_2 y_d(t). \tag{2.241}$$

The constraint equation for $\boldsymbol{x}_d(t)$ becomes a nonlinear differential equation for $z_d(t) = y_d(t)$,

$$\dot{z}_d(t) = R(z_d(t)) - x_d(t). \tag{2.242}$$

To realize the desired output $y_d(t)$, any reference to the inhibitor $x_d(t)$ has to be eliminated from the control signal Eq. (2.241). To achieve that, the constraint equation (2.242) must be solved for $x_d(t)$ in terms of the desired output $z_d(t)$. This is a very simple task because Eq. (2.242) is a linear algebraic equation for $x_d(t)$. The solution is

$$x_d(t) = R(z_d(t)) - \dot{z}_d(t). \tag{2.243}$$

Using the last relation, $x_d(t)$ can be eliminated from the control signal Eq. (2.241) to get

$$\begin{aligned} u(t) &= \dot{x}_d(t) - a_0 - a_1 x_d(t) - a_2 z_d(t) \\ &= R'(z_d(t))\,\dot{z}_d(t) - \ddot{z}_d(t) - a_0 - a_1 z_d(t) - a_2 (R(z_d(t)) - \dot{z}_d(t)). \end{aligned} \tag{2.244}$$

In conclusion, the control signal $u(t)$ as well as the desired state $\boldsymbol{x}_d(t)$ is expressed solely in terms of the desired output $z_d(t)$. Although the system does not satisfy the linearizing assumption because the constraint equation is a nonlinear differential equation, only a linear algebraic equation had to be solved. Thus, linear structures underlying nonlinear control systems may exist independently of the linearizing assumption. Interestingly, the approach of open loop control proposed here yields a similar result for the control as feedback linearization, see e.g. (Khalil, 2001). This hints at deep connections between our approach and feedback linearization. The framework of exactly realizable trajectories might open up a way to generalize feedback linearization to open loop control systems.

A remark about the initial conditions. For exactly realizable trajectories the initial state of the desired trajectory must be equal to the initial system state,

$\boldsymbol{x}_d(t_0) = \boldsymbol{x}(t_0)$. Due to Eq. (2.243), the initial value for x_d is fully determined by the initial value of the desired output $z_d(t_0)$ and its time derivative $\dot{z}_d(t_0)$. For a fixed desired output trajectory $z_d(t)$, the system must be prepared in the initial state

$$x(t_0) = R(z_d(t_0)) - \dot{z}_d(t_0), \tag{2.245}$$
$$y(t_0) = z_d(t_0). \tag{2.246}$$

On the other hand, if the system cannot be prepared in a certain initial state, Eq. (2.243) imposes an additional condition on the desired output trajectory $z_d(t)$. In fact, not only is the initial value $z_d(t_0)$ prescribed by Eq. (2.246), but also the initial value of the time derivative $\dot{z}_d(t_0)$ is fixed by Eq. (2.245).

Figure 2.7 shows the result of a numerical simulation of the controlled FHN model with the prescribed activator

$$y_d(t) = 4\sin(2t) \tag{2.247}$$

as the desired output trajectory. At the initial time $t = t_0 = 0$, the system is prepared in a state such that Eqs. (2.245) and (2.246) are satisfied,

$$\begin{pmatrix} x_0, & y_0 \end{pmatrix}^T = \begin{pmatrix} -8, & 0 \end{pmatrix}^T. \tag{2.248}$$

Numerically solving the controlled system and comparing the controlled state trajectories $\boldsymbol{x}(t)$ with the corresponding desired reference trajectories reveals a perfect agreement within numerical precision, see the bottom row of Fig. 2.7.

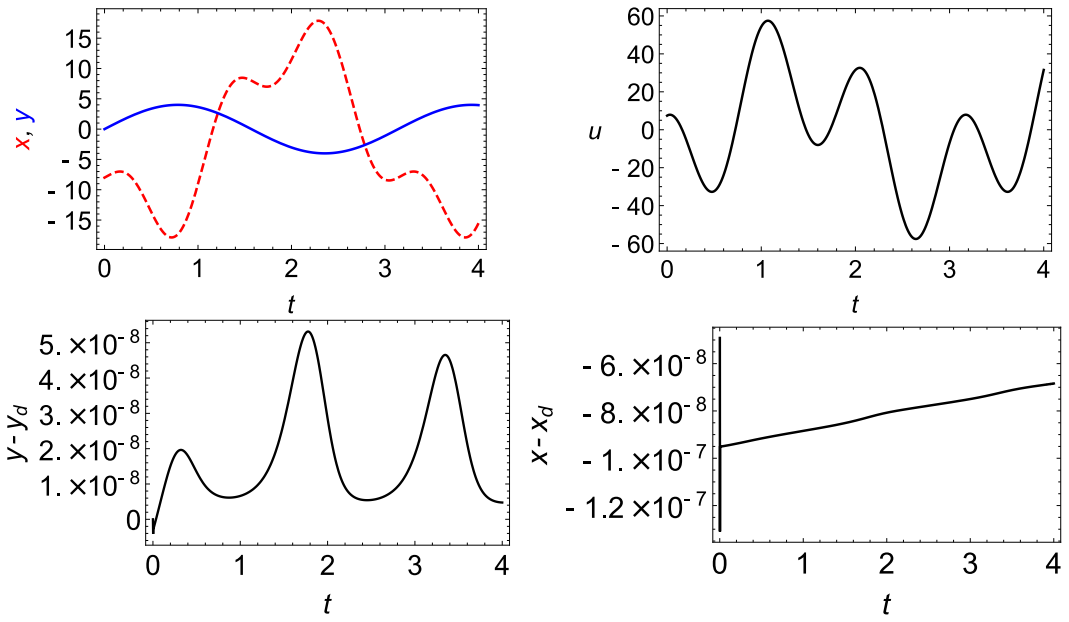

Figure 2.7.: Inhibitor-controlled FHN model with activator y_d as the desired output trajectory. Top left: the desired activator (blue solid line) is prescribed as in Eq. (2.247), while the desired inhibitor x_d (red dashed line) behaves as given by Eq. (2.243). The control u is shown in the top right panel. Comparing the difference between desired activator (bottom left) and inhibitor (bottom right) with the corresponding controlled time evolution obtained from a numerical solution demonstrates agreement within numerical precision.

Example 2.18: Modified Oregonator model

The modified Oregonator model is a model for the light sensitive Belousov-Zhabotinsky reaction (Krug et al., 1990; Field et al., 1972; Field and Noyes, 1974). In experiments, the intensity of illuminated light is used to control the system. The Belousov-Zhabotinsky reaction has been used as an experimental play ground for ideas related to the control of complex systems, see e.g. (Mikhailov and Showalter, 2006) for examples. The system equations for the

activator y and inhibitor x read as

$$\begin{pmatrix} \dot{x}(t) \\ \dot{y}(t) \end{pmatrix} = \begin{pmatrix} y(t) - x(t) \\ \frac{1}{\tilde{\epsilon}} \left(y(t)(1 - y(t)) + f x(t) \frac{q - y(t)}{q + y(t)} \right) \end{pmatrix} + \begin{pmatrix} 0 \\ \frac{1}{\tilde{\epsilon}} \frac{q - y(t)}{q + y(t)} \end{pmatrix} u(t). \tag{2.249}$$

The control signal $u(t)$ is proportional to the applied light intensity. In experiments, the inhibitor is visible and can be recorded with a camera. The measured gray scale depends linearly on the inhibitor and is used as the output z,

$$z(t) = h(x(t)) = I_0 + c x(t). \tag{2.250}$$

For a desired trajectory to be exactly realizable, it has to satisfy the linear constraint equation,

$$\dot{x}_d(t) = y_d(t) - x_d(t). \tag{2.251}$$

Equation (2.251) is solved for $y_d(t)$ to obtain

$$y_d(t) = \dot{x}_d(t) + x_d(t) = \frac{1}{c} \dot{z}_d(t) + \frac{1}{c} (z_d(t) - I_0). \tag{2.252}$$

The inhibitor $x_d(t)$ was substituted with the desired output $z_d(t)$ given by Eq. (2.250). The control signal $u(t)$ can be expressed entirely in terms of the desired output $z_d(t)$ as

$$\begin{aligned} u(t) &= \frac{q + y_d(t)}{q - y_d(t)} \left(\tilde{\epsilon} \dot{y}_d(t) + y_d(t)(y_d(t) - 1) \right) - f x_d(t) \\ &= \frac{\tilde{\epsilon}}{c} \frac{cq + \dot{z}_d(t) + z_d(t) - I_0}{cq - \dot{z}_d(t) - z_d(t) + I_0} \left(\ddot{z}_d(t) + \dot{z}_d(t) \right) \\ &+ \frac{1}{c^2} \frac{cq + \dot{z}_d(t) + z_d(t) - I_0}{cq - \dot{z}_d(t) - z_d(t) + I_0} \left(\dot{z}_d(t) + z_d(t) - I_0 \right) \left(\dot{z}_d(t) + z_d(t) - I_0 - c \right) \\ &- \frac{f}{c} (z_d(t) - I_0). \end{aligned} \tag{2.253}$$

Since only the output $z(t)$ can be observed in experiments, the initial state $\boldsymbol{x}_0(t) = \begin{pmatrix} x_0, & y_0 \end{pmatrix}^T$ of the system must be determined from $z(t)$. Solving Eq.

(2.250) for $x(t)$ and using also Eq. (2.252) yields

$$x_0 = x(t_0) = \frac{1}{c}\left(z(t_0) - I_0\right), \tag{2.254}$$

$$y_0 = y(t_0) = \frac{1}{c}\dot{z}(t_0) + \frac{1}{c}\left(z(t_0) - I_0\right). \tag{2.255}$$

Thus, observation of the full initial state requires knowledge of the output $z(t_0)$ as well as its time derivative $\dot{z}(t_0)$. A generalization of this fact leads to the notion of observability, see e.g. (Chen, 1998) for the definition of observability in the context of linear systems. On the other hand, assuming it is impossible to prepare the system in a desired initial state, the desired output trajectory $z_d(t)$ has to satisfy specific initial conditions to comply with the initial system state $\begin{pmatrix} x_0, & y_0 \end{pmatrix}^T$. In fact, these conditions are identical in form to Eqs. (2.254) and (2.255),

$$x_0 = \frac{1}{c}\left(z_d(t_0) - I_0\right), \tag{2.256}$$

$$y_0 = \frac{1}{c}\dot{z}_d(t_0) + \frac{1}{c}\left(z_d(t_0) - I_0\right). \tag{2.257}$$

In conclusion, for a successful realization of the desired output z_d, not only the initial value of z_d but also its time derivative must be prescribed. This result hints at a connection between output realizability and observability. A similar connection between observability and controllability is known as the principle of duality since the initial work of Kalman (Kalman, 1959), see also (Chen, 1998).

2.7. Conclusions

2.7.1. Summary

A common approach to control, especially in the context of LTI systems, is concerned with states as the objects to be controlled. Suppose a controlled system, often called a plant in this context, has a certain point $\boldsymbol{x}_1$ in state space, sometimes called the operating point, at which the system works efficiently. The control task is then to bring the system to the operating point $\boldsymbol{x}_1$, and keep it there. This naturally leads to a definition of controllability as the possibility to achieve a state-to-state transfer from an initial state $\boldsymbol{x}_0$ to the operating point $\boldsymbol{x}_1$ within finite time (Kalman, 1959; Chen, 1998).

In contrast to that, here an approach to control is developed which centers on the state trajectory $\boldsymbol{x}(t)$ as the object of interest. Of course, both approaches to control are closely related. A single operating point in state space at which the system is to

be kept is nothing more than a degenerate state trajectory. Equivalently, any state trajectory can be approximated by a succession of working points.

We distinguish between the controlled state trajectory $\boldsymbol{x}(t)$ and the desired trajectory $\boldsymbol{x}_d(t)$. The former is the trajectory which the time-dependent state $\boldsymbol{x}$ traces out in state space under the action of a control signal. The latter is a fictitious reference trajectory for the state over time. It is prescribed in analytical or numerical form by the experimenter. Depending on the choice of the desired trajectory $\boldsymbol{x}_d(t)$, the controlled state $\boldsymbol{x}(t)$ may or may not follow $\boldsymbol{x}_d(t)$.

For affine control systems, the class of exactly realizable desired trajectories is defined in Section 2.2. For this subset of desired trajectories, a control signal exists which enforces the controlled state to follow the desired trajectory exactly,

$$\boldsymbol{x}(t) = \boldsymbol{x}_d(t), \tag{2.258}$$

for all times $t \geq t_0$. Exactly realizable desired trajectories satisfy the constraint equation

$$\mathbf{0} = \boldsymbol{\mathcal{Q}}(\boldsymbol{x}_d(t))\left(\dot{\boldsymbol{x}}_d(t) - \boldsymbol{R}(\boldsymbol{x}_d(t))\right), \tag{2.259}$$

with projector

$$\boldsymbol{\mathcal{Q}}(\boldsymbol{x}) = \mathbf{1} - \boldsymbol{\mathcal{B}}(\boldsymbol{x})\,\boldsymbol{\mathcal{B}}^{+}(\boldsymbol{x}) \tag{2.260}$$

with rank $n - p$. The vector of control signals $\boldsymbol{u}(t)$ is expressed in terms of the desired trajectory as

$$\boldsymbol{u}(t) = \boldsymbol{\mathcal{B}}^{+}(\boldsymbol{x}_d(t))\left(\dot{\boldsymbol{x}}_d(t) - \boldsymbol{R}(\boldsymbol{x}_d(t))\right). \tag{2.261}$$

The matrix $\boldsymbol{\mathcal{B}}^{+}(\boldsymbol{x})$ is the Moore-Penrose pseudo inverse of the coupling matrix $\boldsymbol{\mathcal{B}}(\boldsymbol{x})$. Equation (2.261) establishes a one-to-one relationship between the p-dimensional control signal $\boldsymbol{u}(t)$ and p out of n components of the desired trajectory $\boldsymbol{x}_d(t)$. The constraint equation (2.259) fixes those $n - p$ components of the desired trajectory $\boldsymbol{x}_d(t)$ without a one-to-one relationship to the control signal. The projectors $\boldsymbol{\mathcal{P}}(\boldsymbol{x}) = \boldsymbol{\mathcal{B}}(\boldsymbol{x})\,\boldsymbol{\mathcal{B}}^{+}(\boldsymbol{x})$ and $\boldsymbol{\mathcal{Q}}(\boldsymbol{x}) = \mathbf{1} - \boldsymbol{\mathcal{P}}(\boldsymbol{x})$ allow a coordinate-free separation of the state $\boldsymbol{x}$ as well as the controlled state equation in two parts. The part of the state equation proportional to $\boldsymbol{\mathcal{P}}(\boldsymbol{x})$ determines the control signal. This approach allows the elimination of the control signal Eq. (2.261) from the system. The remaining part of the state equation, which is proportional to $\boldsymbol{\mathcal{Q}}(\boldsymbol{x})$, is the constraint equation (2.259). For the control of exactly realizable trajectories, only the constraint equation must be solved.

Note that the control signal Eq. (2.261) does not depend on the state of the system and is therefore an open loop control. As such, it may suffer from instability. An exactly realizable desired trajectory might or might not be stable against perturbations of the initial conditions or external perturbations as e.g. noise.

On the basis of the control signal Eq. (2.261) and constraint equation (2.259), a hierarchy of desired trajectories $\boldsymbol{x}_d(t)$ comprising 3 classes is established:

(A) desired trajectories $\boldsymbol{x}_d(t)$ which are solutions to the uncontrolled system,

(B) desired trajectories $\boldsymbol{x}_d(t)$ which are exactly realizable,

(C) arbitrary desired trajectories $\boldsymbol{x}_d(t)$.

Desired trajectories of class (A) satisfy the uncontrolled state equation

$$\dot{\boldsymbol{x}}_d(t) = \boldsymbol{R}(\boldsymbol{x}_d(t)). \tag{2.262}$$

This constitutes the most specific class of desired trajectories. Because of Eq. (2.262), the constraint equation (2.259) is trivially satisfied and the control signal as given by Eq. (2.261) vanishes,

$$\boldsymbol{u}(t) = \boldsymbol{0}. \tag{2.263}$$

Equation (2.263) implies a non-invasive control signal, i.e., the control signal vanishes upon achieving the control target. Because of Eq. (2.263), the open loop control approach proposed here cannot be employed for desired trajectories of class (A). Instead, these desired trajectories require feedback control. Class (A) encompasses several important control tasks, as e.g. the stabilization of unstable stationary states (Sontag, 2011). A prominent example extensively studied by the physics community is the control of chaotic systems by small perturbations (Ott et al., 1990; Shinbrot et al., 1993). One of the fundamental aspects of chaos is that many different possible motions are simultaneously present in the system. In particular, an infinite number of unstable periodic orbits co-exist with the chaotic motion. All orbits are solutions to the uncontrolled system dynamics Eq. (2.262). Using non-invasive feedback control, a particular orbit may be stabilized. See also (Schöll and Schuster, 2007; Schimansky-Geier et al., 2007) and references therein for more information and examples.

Desired trajectories of class (B) satisfy the constraint equation (2.259) and yield a non-vanishing control signal $\boldsymbol{u}(t) \neq \boldsymbol{0}$. The approach developed in this chapter applies to this class. Several other techniques developed in mathematical control theory, as e.g. feedback linearization and differential flatness, also work with this class of desired trajectories (Khalil, 2001; Sira-Ramírez and Agrawal, 2004). Class (B) contains the desired trajectories from class (A) as a special case. For desired trajectories of class (A) and class (B), the solution of the controlled state trajectory is simply given by $\boldsymbol{x}(t) = \boldsymbol{x}_d(t)$.

Finally, class (C) is the most general class of desired trajectories and contains class (A) and (B) as special cases. In general, these desired trajectories do not satisfy the constraint equation,

$$\boldsymbol{0} \neq \boldsymbol{\mathcal{Q}}(\boldsymbol{x}_d(t))\left(\dot{\boldsymbol{x}}_d(t) - \boldsymbol{R}(\boldsymbol{x}_d(t))\right), \tag{2.264}$$

such that, in general, the approach developed in this chapter cannot be applied to desired trajectories of class (C). No general expression for the control signal in terms

of the desired trajectory $\boldsymbol{x}_d(t)$ is available. In general, the solution for the controlled state trajectory $\boldsymbol{x}(t)$ is not simply given by $\boldsymbol{x}_d(t)$, $\boldsymbol{x}(t) \neq \boldsymbol{x}_d(t)$. Thus, a solution to control problems defined by class (C) does not only consist in finding an expression for the control signal, but also involves finding a solution for the controlled state trajectory $\boldsymbol{x}(t)$ as well. One possible method to solve such control problems is optimal control.

The linearizing assumption of Section 2.3 defines a class of nonlinear control systems which essentially behave like linear control system. Models satisfying the linearizing assumption allow exact analytical solutions in closed form even if no analytical solutions for the uncontrolled system exists, see e.g. the SIR model in Example 2.16. The linearizing assumption uncovers a hidden linear structure underlying nonlinear control systems. Similarly, feedback linearization defines a huge class of nonlinear control systems possessing an underlying linear structure. The class of feedback linearizable systems contains the systems satisfying the linearizing assumption as a trivial case. However, the linearizing assumption defined here goes much further than feedback linearization. In fact, while general nonlinear control systems require a fairly abstract treatment for the definition of controllability (Slotine and Li, 1991; Isidori, 1995), we were able to apply the relatively simple notion of controllability in terms of a rank condition to systems satisfying the linearizing assumption, see Section 2.4. This is a direct extension of the properties of linear control systems to a class of nonlinear control systems. Furthermore, as will be shown in the next two chapters, the class defined by the linearizing assumption exhibits a linear structure even in case of optimal control for arbitrary, not necessarily exactly realizable desired trajectories. This enables the determination of exact, closed form expressions for optimal trajectory tracking in Chapter 4.

The approach to control proposed here shares many similarities to theories developed in mathematical control theory. We already mentioned inverse dynamics in the context of mechanical systems in Example 2.4. For more information about inverse dynamics, we refer the reader to the literature about robot control (Lewis et al., 1993; de Wit et al., 2012; Angeles, 2013). In the following, we analyze the similarities and differences of our approach with differential flatness.

2.7.2. Differential flatness

Similar to the concept of exactly realizable trajectories proposed in this chapter, differential flatness provides an open loop method for the control of dynamical systems. We first give a short introduction to differential flatness to be able to compare the similarities and differences to our approach. For more information about differential flatness as well as many examples, we refer the reader to (Fliess et al., 1995; Van Nieuwstadt and Murray, 1997; Sira-Ramírez and Agrawal, 2004; Levine, 2009). The presentation follows (Sira-Ramírez and Agrawal, 2004).

Differential flatness relies on the notion of differential functions. A function $\boldsymbol{\phi}$ is

a differential function of $\boldsymbol{x}(t)$ if it depends on $\boldsymbol{x}(t)$ and its time derivatives up to order β,

$$\boldsymbol{\phi}(t) = \boldsymbol{\phi}\left(\boldsymbol{x}(t), \dot{\boldsymbol{x}}(t), \ldots, \boldsymbol{x}^{(\beta)}(t)\right). \tag{2.265}$$

The symbol

$$\boldsymbol{x}^{(\beta)}(t) = \frac{d^{\beta}}{dt^{\beta}}\boldsymbol{x}(t) \tag{2.266}$$

denotes the time derivative of order β. An affine control system with n-component state vector $\boldsymbol{x}$ and p-component control signal $\boldsymbol{u}$ satisfies

$$\dot{\boldsymbol{x}}(t) = \boldsymbol{R}(\boldsymbol{x}(t)) + \boldsymbol{\mathcal{B}}(\boldsymbol{x}(t))\,\boldsymbol{u}(t). \tag{2.267}$$

Applying differentiation with respect to time to Eq. (2.267), the differential function $\tilde{\boldsymbol{\phi}}(\boldsymbol{x}, \dot{\boldsymbol{x}})$ can be expressed as a function of $\boldsymbol{x}$ and $\boldsymbol{u}$

$$\boldsymbol{\phi}(\boldsymbol{x}(t), \boldsymbol{u}(t)) = \tilde{\boldsymbol{\phi}}(\boldsymbol{x}(t), \dot{\boldsymbol{x}}(t)) = \tilde{\boldsymbol{\phi}}(\boldsymbol{x}(t), \boldsymbol{R}(\boldsymbol{x}(t)) + \boldsymbol{\mathcal{B}}(\boldsymbol{x}(t))\,\boldsymbol{u}(t)). \tag{2.268}$$

Similarly, the differential function $\tilde{\boldsymbol{\phi}}\left(\boldsymbol{x}(t), \dot{\boldsymbol{x}}(t), \ldots, \boldsymbol{x}^{(\beta)}(t)\right)$ can be expressed as a differential function $\boldsymbol{\phi}\left(\boldsymbol{x}(t), \boldsymbol{u}(t), \dot{\boldsymbol{u}}(t), \ldots, \boldsymbol{u}^{(\beta-1)}(t)\right)$.

The system Eq. (2.267) is called differentially flat if there exists a p-component fictional output $\boldsymbol{z}(t) = \begin{pmatrix} z_1(t), & \ldots, & z_p(t) \end{pmatrix}^T$ such that (Sira-Ramírez and Agrawal, 2004)

1. the output $\boldsymbol{z}(t)$ is representable as a differential function of the state $\boldsymbol{x}(t)$ and the vector of control signals $\boldsymbol{u}(t)$ as

$$\boldsymbol{z}(t) = \boldsymbol{\phi}\left(\boldsymbol{x}(t), \boldsymbol{u}(t), \dot{\boldsymbol{u}}(t), \ddot{\boldsymbol{u}}(t), \ldots, \boldsymbol{u}^{(\beta-1)}(t)\right), \tag{2.269}$$

2. the state $\boldsymbol{x}(t)$ and the vector of control signals $\boldsymbol{u}(t)$ are representable as a differential function of the output $\boldsymbol{z}(t)$ as (with finite integer α)

$$\boldsymbol{x}(t) = \boldsymbol{\chi}\left(\boldsymbol{z}(t), \dot{\boldsymbol{z}}(t), \ddot{\boldsymbol{z}}(t), \ldots, \boldsymbol{z}^{(\alpha)}(t)\right), \tag{2.270}$$

$$\boldsymbol{u}(t) = \boldsymbol{\psi}\left(\boldsymbol{z}(t), \dot{\boldsymbol{z}}(t), \ddot{\boldsymbol{z}}(t), \ldots, \boldsymbol{z}^{(\alpha+1)}(t)\right), \tag{2.271}$$

3. the components of the output $\boldsymbol{z}(t)$ are differentially independent, i.e., they satisfy no differential equation of the form

$$\boldsymbol{\Omega}\left(\boldsymbol{z}(t), \dot{\boldsymbol{z}}(t), \ddot{\boldsymbol{z}}(t), \ldots, \boldsymbol{z}^{(\beta)}(t)\right) = \boldsymbol{0}. \tag{2.272}$$

For a differentially flat system, the full solution for the state over time $\boldsymbol{x}(t)$ as well as the control signal $\boldsymbol{u}(t)$ can be expressed in terms of the output over time $\boldsymbol{z}(t)$.

Mathematically, the relation between control signal $\boldsymbol{u}(t)$ and output $\boldsymbol{z}(t)$ is a differential function. This has the great advantage that the determination of the control signal can be done in real time at time t by computing only a finite number of time derivatives of $\boldsymbol{z}(t)$. This would not be possible if $\boldsymbol{u}(t)$ also involves time integrals of the output $\boldsymbol{z}$ because these would require summation over all previous times as well. Another advantage is that no differential equations need to be solved to obtain the control signal and state trajectory. Usually, all expressions are generated by simply differentiating the controlled state equations with respect to time. The output $\boldsymbol{z}(t)$ has the same number p of components as the number of independent input signals $\boldsymbol{u}(t)$ available to control the system. If the control signal determined by $\boldsymbol{\psi}$, Eq. (2.271), is applied to the system, then the system's output is $\boldsymbol{z}(t)$. Differentially flat systems are not necessarily affine in control but can be nonlinear in the control as well. However, only certain systems are differentially flat, and it is not known under which conditions a controlled dynamical system is differentially flat if the number of independent control signals is larger than one, $p > 1$.

Similar to differential flatness, this chapter proposes an open loop control method. A solution for control signals exactly realizing a desired output $\boldsymbol{z}_d(t)$ with p components is determined. In the discussion of output trajectory realizability in Section 2.6, the control signal is expressed solely in terms of the desired output $\boldsymbol{z}_d(t)$ and the initial conditions for the state. This implies that the controlled state trajectory, given as the solution to the controlled state equation, can also be expressed in terms of the desired output and the initial conditions for the state. These facts fully agree with the concept of differential flatness. In contrast to the approach here, the literature about differential flatness does usually not distinguish explicitly between desired trajectory $\boldsymbol{x}_d(t)$ and controlled state trajectory $\boldsymbol{x}(t)$, but implicitly assumes this identity from the very beginning.

The most striking difference between the approach here and differential flatness is the restriction to differential functions. In general, our approach yields a control signal in terms of a functional of the desired output,

$$\boldsymbol{u}(t) = \boldsymbol{u}\left[\boldsymbol{z}_d(t)\right]. \tag{2.273}$$

Note that a functional is a more general expression than a differential function. Using the Dirac delta function $\delta(t)$, any time derivative of order β can be expressed as a functional,

$$\begin{aligned}
\boldsymbol{z}_d^{(\beta)}(t) &= \int_{-\infty}^{\infty} d\tau \delta(\tau - t)\, \boldsymbol{z}_d^{(\beta)}(\tau) = -\int_{-\infty}^{\infty} d\tau \delta'(\tau - t)\, \boldsymbol{z}_d^{(\beta-1)}(\tau) \\
&\vdots \\
&= (-1)^{\beta} \int_{-\infty}^{\infty} d\tau \delta^{(\beta)}(\tau - t)\, \boldsymbol{z}_d(\tau)
\end{aligned} \tag{2.274}$$

Therefore, any differential function of $\boldsymbol{z}(t)$ can be expressed in terms of a function of functionals of $\boldsymbol{z}(t)$, while the reverse is not true. The restriction to differential functions might also explain why only certain systems are differentially flat. In contrast, the approach proposed here can be applied to any affine control system. As an advantage, differential flatness yields expressions for state and control which are computationally more efficient because they do not require the solution of differential equations or integrals, which is not the case here.

2.7.3. Outlook

The framework of exactly realizable trajectories is interpreted as an open loop control method. However, it may be possible to extend this approach to feedback control. As discussed in Section 2.6, a control $\boldsymbol{u}(t)$ realizing a p-component desired output $\boldsymbol{z}_d(t) = \boldsymbol{h}(\boldsymbol{x}_d(t))$ is expressed entirely in terms of the desired output. The dependence of $\boldsymbol{u}(t)$ on $\boldsymbol{z}_d(t)$ is typically in form of a functional,

$$\boldsymbol{u}(t) = \boldsymbol{u}[\boldsymbol{z}_d(t)]. \tag{2.275}$$

A generalization to feedback control yields a control signal which does not only depend on the desired output $\boldsymbol{z}_d(t)$ but also on the monitored state $\boldsymbol{x}(t)$ of the controlled system,

$$\boldsymbol{u}(t) = \boldsymbol{u}[\boldsymbol{z}_d(t), \boldsymbol{x}(t)]. \tag{2.276}$$

In general, the control signal is allowed to depend on the history of $\boldsymbol{x}(t)$ such that the dependence of $\boldsymbol{u}(t)$ on $\boldsymbol{x}(t)$ is also in form of a functional. Such a generalization of the approach to control proposed here certainly changes the stability properties of the controlled system and may result in an improved stability of the controlled trajectory.

A fundamental problem affecting not only exactly realizable trajectories but also feedback linearization, differential flatness, and optimal control, is the requirement of exactly knowing the system dynamics. This must be contrasted with the fact that the majority of physical models are idealizations. Unknown external influences in control systems can be modeled as noise or structural perturbations, which might both depend on the system state itself. To ensure a successful control in experiments, the proposed control methods must not only be stable against perturbations of the initial conditions, but must be sufficiently stable against structural perturbations as well. Stability against structural perturbations is also known as robustness in the context of control theory (Freeman and Kokotovic, 1996). Before applying the control method developed in this chapter to real world problems, a thorough investigation of the stability of the control problem at hand must be conducted. In case of instability, countermeasures as e.g. additional stabilizing feedback control must be applied (Khalil, 2001).

Section 2.3 introduces the linearizing assumption. On the one hand, this assumption is restrictive, but on the other hand it has far reaching consequences and results in significant simplifications for nonlinear affine control systems. A possible generalization of the linearizing assumption might be as follows. First, relax condition Eq. (2.80) and allow a state dependent projector $\boldsymbol{\mathcal{Q}}(\boldsymbol{x})$ which, however, does only depend on the state components $\boldsymbol{\mathcal{P}}\boldsymbol{x}$,

$$\boldsymbol{\mathcal{Q}}(\boldsymbol{x}) = \boldsymbol{\mathcal{Q}}(\boldsymbol{\mathcal{P}}\boldsymbol{x} + \boldsymbol{\mathcal{Q}}\boldsymbol{x}) = \boldsymbol{\mathcal{Q}}(\boldsymbol{\mathcal{P}}\boldsymbol{x}). \tag{2.277}$$

Second, also relax condition Eq. (2.82) and assume a nonlinearity $\boldsymbol{R}(\boldsymbol{x})$ with the following structure,

$$\boldsymbol{\mathcal{Q}}(\boldsymbol{x})\,\boldsymbol{R}(\boldsymbol{x}) = \boldsymbol{\mathcal{Q}}(\boldsymbol{\mathcal{P}}\boldsymbol{x})\,\boldsymbol{\mathcal{A}}(\boldsymbol{\mathcal{P}}\boldsymbol{x})\,\boldsymbol{\mathcal{Q}}(\boldsymbol{\mathcal{P}}\boldsymbol{x})\,\boldsymbol{x} + \boldsymbol{\mathcal{Q}}(\boldsymbol{\mathcal{P}}\boldsymbol{x})\,\boldsymbol{b}(\boldsymbol{\mathcal{P}}\boldsymbol{x}). \tag{2.278}$$

The matrix $\boldsymbol{\mathcal{A}}(\boldsymbol{\mathcal{P}}\boldsymbol{x})$, the projector $\boldsymbol{\mathcal{Q}}(\boldsymbol{\mathcal{P}}\boldsymbol{x})$ and the inhomogeneity $\boldsymbol{b}(\boldsymbol{\mathcal{P}}\boldsymbol{x})$ may all depend on the state components $\boldsymbol{\mathcal{P}}\boldsymbol{x}$.

Together with

$$\begin{aligned}\frac{d}{dt}\left(\boldsymbol{\mathcal{Q}}(\boldsymbol{\mathcal{P}}\boldsymbol{x}_d)\,\boldsymbol{x}_d\right) = {}& \dot{\boldsymbol{\mathcal{Q}}}(\boldsymbol{\mathcal{P}}\boldsymbol{x}_d)\,\boldsymbol{\mathcal{P}}(\boldsymbol{\mathcal{P}}\boldsymbol{x}_d)\,\boldsymbol{x}_d \\ & + \dot{\boldsymbol{\mathcal{Q}}}(\boldsymbol{\mathcal{P}}\boldsymbol{x}_d)\,\boldsymbol{\mathcal{Q}}(\boldsymbol{\mathcal{P}}\boldsymbol{x}_d)\,\boldsymbol{x}_d + \boldsymbol{\mathcal{Q}}(\boldsymbol{\mathcal{P}}\boldsymbol{x}_d)\,\dot{\boldsymbol{x}}_d,\end{aligned} \tag{2.279}$$

the constraint equation becomes

$$\frac{d}{dt}\left(\boldsymbol{\mathcal{Q}}\boldsymbol{x}_d(t)\right) = \left(\dot{\boldsymbol{\mathcal{Q}}} + \boldsymbol{\mathcal{Q}}\boldsymbol{\mathcal{A}}\right)\boldsymbol{\mathcal{Q}}\boldsymbol{x}_d(t) + \dot{\boldsymbol{\mathcal{Q}}}\boldsymbol{\mathcal{P}}\boldsymbol{x}_d(t) + \boldsymbol{\mathcal{Q}}\boldsymbol{b}. \tag{2.280}$$

The arguments are suppressed and it is understood that $\dot{\boldsymbol{\mathcal{Q}}}$, $\boldsymbol{\mathcal{Q}}$, $\boldsymbol{\mathcal{P}}$, $\boldsymbol{\mathcal{A}}$, and $\boldsymbol{b}$ may depend on the part $\boldsymbol{\mathcal{P}}\boldsymbol{x}_d(t)$. Equation (2.280) is a linear equation for $\boldsymbol{\mathcal{Q}}\boldsymbol{x}_d(t)$ and can thus be solved with the help of its state transition matrix, see Appendix A.1. However, the matrix $\boldsymbol{\mathcal{A}} = \boldsymbol{\mathcal{A}}(\boldsymbol{\mathcal{P}}\boldsymbol{x}_d(t))$ exhibits an explicit time dependence through its dependence on $\boldsymbol{\mathcal{P}}\boldsymbol{x}_d(t)$. This necessitates modifications for the notion of controllability from Section 2.4, see also (Chen, 1998).

A central assumption of the formalism presented in this chapter is that the $n \times p$ coupling matrix $\boldsymbol{\mathcal{B}}(\boldsymbol{x})$ has full rank p for all values of $\boldsymbol{x}$. This assumption leads to a Moore-Penrose pseudo inverse $\boldsymbol{\mathcal{B}}^{+}(\boldsymbol{x})$ of $\boldsymbol{\mathcal{B}}(\boldsymbol{x})$ given by

$$\boldsymbol{\mathcal{B}}^{+}(\boldsymbol{x}) = \left(\boldsymbol{\mathcal{B}}^{T}(\boldsymbol{x})\,\boldsymbol{\mathcal{B}}(\boldsymbol{x})\right)^{-1}\boldsymbol{\mathcal{B}}^{T}(\boldsymbol{x}). \tag{2.281}$$

If $\boldsymbol{\mathcal{B}}(\boldsymbol{x})$ does not have full rank for some or all values of $\boldsymbol{x}$, the inverse of $\boldsymbol{\mathcal{B}}^{T}(\boldsymbol{x})\,\boldsymbol{\mathcal{B}}(\boldsymbol{x})$ does not exist. However, a unique Moore-Penrose pseudo inverse $\boldsymbol{\mathcal{B}}^{+}(\boldsymbol{x})$ does exist for any matrix $\boldsymbol{\mathcal{B}}(\boldsymbol{x})$, regardless of its rank. No closed form expressions exist for the general case, but $\boldsymbol{\mathcal{B}}^{+}(\boldsymbol{x})$ can nevertheless be computed numerically by singular value decomposition, for example. Because $\boldsymbol{\mathcal{B}}^{+}(\boldsymbol{x})$ exists in any case, the $n \times n$ projector defined by

$$\boldsymbol{\mathcal{P}}(\boldsymbol{x}) = \boldsymbol{\mathcal{B}}(\boldsymbol{x})\,\boldsymbol{\mathcal{B}}^{+}(\boldsymbol{x}) \tag{2.282}$$

exists as well. Thus, using the general Moore-Penrose pseudo inverse $\boldsymbol{\mathcal{B}}^{+}(\boldsymbol{x})$, the formalism developed in this chapter can be extended to cases with $\boldsymbol{\mathcal{B}}(\boldsymbol{x})$ not having full rank for some values of $\boldsymbol{x}$.

A mathematically more rigorous treatment of the notion of exactly realizable trajectories is desirable. An important question is the following. Under which conditions does the constraint equation

$$\mathbf{0} = \boldsymbol{\mathcal{Q}}(\boldsymbol{x}_d(t))(\dot{\boldsymbol{x}}_d(t) - \boldsymbol{R}(\boldsymbol{x}_d(t))) \tag{2.283}$$

have a unique solution for $\boldsymbol{\mathcal{Q}}(\boldsymbol{x}_d(t))\boldsymbol{x}_d(t)$? Note that Eq. (2.283) is a non-autonomous nonlinear system of differential equations for $\boldsymbol{\mathcal{Q}}(\boldsymbol{x}_d(t))\boldsymbol{x}_d(t)$ with the explicit time dependence caused by the part $\boldsymbol{\mathcal{P}}(\boldsymbol{x}_d(t))\boldsymbol{x}_d(t)$. Therefore, a related question is for conditions on the part $\boldsymbol{\mathcal{P}}(\boldsymbol{x}_d(t))\boldsymbol{x}_d(t)$ prescribed by the experimenter. For example, is $\boldsymbol{\mathcal{P}}(\boldsymbol{x}_d(t))\boldsymbol{x}_d(t)$ required to be a continuously differentiable function or is it allowed to have jumps? Although some general answers might be possible, such questions are simpler to answer for specific control systems.

3. Optimal control

This chapter introduces the standard approach to optimal control theory in form of the necessary optimality conditions in Section 3.1. Additional necessary optimality conditions for singular optimal control problems, the so-called Kelly or generalized Legendre-Clebsch conditions, are presented in Section 3.2. Section 3.3 gives a brief discussion of the difficulties involved in finding a numerical solution to an optimal control problem. The conditions under which the control of exactly realizable trajectories is optimal are clarified in Section 3.4. The last Section 3.5 presents a simple linear optimal control problem for which an exact but cumbersome analytical solution can be derived. Assuming a small regularization parameter $0 < \epsilon \ll 1$, the exact solution is approximated and a simpler expression is obtained. Additionally, the impact of different terminal conditions on the solution is investigated.

3.1. The necessary optimality conditions

The foundations of optimal control theory were laid in the 1950's and early 1960's. The Russian school of Lev Pontryagin and his students developed the minimum principle (Pontryagin and Boltyanskii, 1962); also called the maximum principle in the Russian literature. This principle is based on the calculus of variations and contains the Euler-Lagrange equations of uncontrolled dynamical systems as a special case. An American school, led by Richard Bellmann, developed the Dynamical Programming approach (Bellman, 2003) based on partial differential equations (PDEs). Both approaches treat essentially the same problem and yield equivalent results. Which approach is preferred is, to some extent, a matter of taste. Here, optimal control is discussed in the framework of Pontryagin's minimum optimal control. An excellent introduction for both approaches to optimal control is the book by Bryson and Ho (Bryson and Ho, 1975). A more elementary and technical approach to the calculus of variations and optimal control is provided by the readable introduction (Liberzon, 2012). See also (Hull, 2003) for applications of optimal control and the mathematically rigorous treatment (Vinter, 2000).

3.1.1. Statement of the problem

Optimal control is concerned with minimizing a target functional $\mathcal{J}$, also called the performance index,

$$\mathcal{J}\left[\boldsymbol{x}\left(t\right),\boldsymbol{u}\left(t\right)\right]=\int_{t_0}^{t_1}dtL\left(\boldsymbol{x}\left(t\right),t\right)+M\left(\boldsymbol{x}\left(t_1\right),t_1\right)+\frac{\epsilon^2}{2}\int_{t_0}^{t_1}dt\left(\boldsymbol{u}\left(t\right)\right)^2. \tag{3.1}$$

Here, $L\left(\boldsymbol{x},t\right)$ is the cost function and $M\left(\boldsymbol{x},t\right)$ is the terminal cost. The target functional Eq. (3.1) is to be minimized subject to the constraint that $\boldsymbol{x}\left(t\right)$ is given as the solution to the controlled dynamical system

$$\dot{\boldsymbol{x}}\left(t\right)=\boldsymbol{R}\left(\boldsymbol{x}\left(t\right)\right)+\boldsymbol{\mathcal{B}}\left(\boldsymbol{x}\left(t\right)\right)\boldsymbol{u}\left(t\right), \tag{3.2}$$

with initial condition

$$\boldsymbol{x}\left(t_0\right)=\boldsymbol{x}_0. \tag{3.3}$$

The parameter ϵ is called the regularization parameter. Furthermore, the state $\boldsymbol{x}$ satisfies the $q\leq n$ end point conditions

$$\boldsymbol{\psi}\left(\boldsymbol{x}\left(t_1\right)\right)=\left(\psi_1\left(\boldsymbol{x}\left(t_1\right)\right),\ldots,\psi_q\left(\boldsymbol{x}\left(t_1\right)\right)\right)^T=\boldsymbol{0}. \tag{3.4}$$

3.1.2. Derivation of the necessary optimality conditions

Following a standard procedure (Bryson and Ho, 1975), the constrained optimization problem is converted to an unconstrained optimization problem. Similar to minimizing an ordinary function under constraints, this is done by introducing Lagrange multipliers. However, in optimal control, the constraint is a differential equation defined on a certain time interval. Consequently, the Lagrange multipliers are functions of time and denoted by $\boldsymbol{\lambda}\left(t\right)=\left(\lambda_1\left(t\right),\ldots,\lambda_n\left(t\right)\right)^T\in\mathbb{R}^n$. The vector $\boldsymbol{\lambda}\left(t\right)$ is called the *co-state* or adjoint state. To accommodate the end point condition $\boldsymbol{\psi}$, additional constant Lagrange multipliers $\boldsymbol{\nu}=\left(\nu_1,\ldots,\nu_q\right)^T\in\mathbb{R}^q$ are introduced. The constrained optimization problem is reduced to the minimization of the unconstrained functional

$$\begin{aligned}\bar{\mathcal{J}}\left[\boldsymbol{x}\left(t\right),\boldsymbol{u}\left(t\right),\boldsymbol{\lambda}\left(t\right),\boldsymbol{\nu}\right]=&\mathcal{J}\left[\boldsymbol{x}\left(t\right),\boldsymbol{u}\left(t\right)\right]+\boldsymbol{\nu}^T\boldsymbol{\psi}\left(\boldsymbol{x}\left(t_1\right)\right)\\&+\int_{t_0}^{t_1}dt\boldsymbol{\lambda}^T\left(t\right)\left(\boldsymbol{R}\left(\boldsymbol{x}\left(t\right)\right)+\boldsymbol{\mathcal{B}}\left(\boldsymbol{x}\left(t\right)\right)\boldsymbol{u}\left(t\right)-\dot{\boldsymbol{x}}\left(t\right)\right).\end{aligned} \tag{3.5}$$

Introducing the control Hamiltonian

$$\begin{aligned}H\left(\boldsymbol{x}\left(t\right),\boldsymbol{u}\left(t\right),t\right)=&L\left(\boldsymbol{x}\left(t\right),t\right)+\frac{\epsilon^2}{2}\left(\boldsymbol{u}\left(t\right)\right)^2\\&+\boldsymbol{\lambda}^T\left(t\right)\left(\boldsymbol{R}\left(\boldsymbol{x}\left(t\right)\right)+\boldsymbol{\mathcal{B}}\left(\boldsymbol{x}\left(t\right)\right)\boldsymbol{u}\left(t\right)\right),\end{aligned} \tag{3.6}$$

and applying partial integration for the term involving $\boldsymbol{\lambda}^T(t)\,\dot{\boldsymbol{x}}(t)$ in Eq. (3.5) yields

$$\bar{\mathcal{J}}\left[\boldsymbol{x}(t),\boldsymbol{u}(t),\boldsymbol{\lambda}(t),\boldsymbol{\nu}\right] = \int_{t_0}^{t_1} dt H\left(\boldsymbol{x}(t),\boldsymbol{u}(t),t\right) + M\left(\boldsymbol{x}(t_1),t_1\right) + \boldsymbol{\nu}^T\boldsymbol{\psi}\left(\boldsymbol{x}(t_1)\right) - \boldsymbol{\lambda}^T(t_1)\,\boldsymbol{x}(t_1) + \boldsymbol{\lambda}^T(t_0)\,\boldsymbol{x}(t_0) + \int_{t_0}^{t_1} dt\dot{\boldsymbol{\lambda}}^T(t)\,\boldsymbol{x}(t). \quad (3.7)$$

The functional $\bar{\mathcal{J}}\left[\boldsymbol{x}(t),\boldsymbol{u}(t),\boldsymbol{\lambda}(t),\boldsymbol{\nu}\right]$ must be minimized with respect to $\boldsymbol{\lambda}(t)$, $\boldsymbol{\lambda}(t_1)$, $\boldsymbol{u}(t)$, $\boldsymbol{\nu}$, and $\boldsymbol{x}(t)$. The initial condition $\boldsymbol{x}_0$ is prescribed and is therefore kept fixed. For $\bar{\mathcal{J}}$ to be extremal, it has to satisfy the variational equations

$$\frac{\delta\bar{\mathcal{J}}}{\delta\boldsymbol{x}(t)} = \boldsymbol{0}, \qquad \frac{\delta\bar{\mathcal{J}}}{\delta\boldsymbol{u}(t)} = \boldsymbol{0}, \qquad \frac{\delta\bar{\mathcal{J}}}{\delta\boldsymbol{\lambda}(t)} = \boldsymbol{0}, \qquad \frac{\delta\bar{\mathcal{J}}}{\delta\boldsymbol{\nu}} = \boldsymbol{0}, \qquad \frac{\delta\bar{\mathcal{J}}}{\delta\boldsymbol{\lambda}(t_1)} = \boldsymbol{0}. \quad (3.8)$$

The variation $\frac{\delta\bar{\mathcal{J}}}{\delta\boldsymbol{u}(t)} = \boldsymbol{0}$ with respect to the vector of control signals $\boldsymbol{u}$ leads to p algebraic equations for the control vector $\boldsymbol{u}(t)$,

$$\epsilon^2\boldsymbol{u}^T(t) + \boldsymbol{\lambda}^T(t)\,\boldsymbol{\mathcal{B}}\left(\boldsymbol{x}(t)\right) = \boldsymbol{0}. \quad (3.9)$$

Transposing yields

$$\epsilon^2\boldsymbol{u}(t) + \boldsymbol{\mathcal{B}}^T\left(\boldsymbol{x}(t)\right)\boldsymbol{\lambda}(t) = \boldsymbol{0}. \quad (3.10)$$

The variation $\frac{\delta\bar{\mathcal{J}}}{\delta\boldsymbol{\lambda}^T(t)} = \boldsymbol{0}$ with respect to the co-state $\boldsymbol{\lambda}(t)$ leads to the controlled state equation

$$\dot{\boldsymbol{x}}(t) = \boldsymbol{R}\left(\boldsymbol{x}(t)\right) + \boldsymbol{\mathcal{B}}\left(\boldsymbol{x}(t)\right)\boldsymbol{u}(t), \quad (3.11)$$

$$\boldsymbol{x}(t_0) = \boldsymbol{x}_0. \quad (3.12)$$

The variation $\frac{\delta\bar{\mathcal{J}}}{\delta\boldsymbol{x}(t)} = \boldsymbol{0}$ with respect to the state $\boldsymbol{x}(t)$ leads to the so-called *adjoint* or *co-state equation*

$$-\dot{\boldsymbol{\lambda}}^T(t) = \boldsymbol{\lambda}^T(t)\,\nabla\boldsymbol{R}\left(\boldsymbol{x}(t)\right) + \boldsymbol{\lambda}^T(t)\,\nabla\boldsymbol{\mathcal{B}}\left(\boldsymbol{x}(t)\right)\boldsymbol{u}(t) + \nabla L\left(\boldsymbol{x}(t),t\right). \quad (3.13)$$

Here, the $n \times n$ Jacobi matrix of the n-dimensional nonlinear function $\boldsymbol{R}$ is defined as

$$\left(\nabla\boldsymbol{R}\left(\boldsymbol{x}\right)\right)_{ij} = \frac{\partial}{\partial x_j}R_i\left(\boldsymbol{x}\right). \quad (3.14)$$

The $n \times p \times n$ Jacobi matrix of the $n \times p$ matrix $\boldsymbol{\mathcal{B}}(\boldsymbol{x})$ is given by

$$\left(\nabla\boldsymbol{\mathcal{B}}\left(\boldsymbol{x}\right)\right)_{ijk} = \frac{\partial}{\partial x_k}\mathcal{B}_{ij}\left(\boldsymbol{x}\right). \quad (3.15)$$

Written component-wise, the inner product of $\boldsymbol{\lambda}$ and $\boldsymbol{\mathcal{B}}(\boldsymbol{x})\boldsymbol{u}$ is

$$\boldsymbol{\lambda}^T \boldsymbol{\mathcal{B}}(\boldsymbol{x})\boldsymbol{u} = \sum_{i=1}^{n}\sum_{j=1}^{p} \lambda_i \mathcal{B}_{ij}(\boldsymbol{x})\, u_j. \tag{3.16}$$

Consequently, the expression $\boldsymbol{\lambda}^T \nabla \boldsymbol{\mathcal{B}}(\boldsymbol{x})\boldsymbol{u}$ is an n-component row vector defined as

$$\boldsymbol{\lambda}^T \nabla \boldsymbol{\mathcal{B}}(\boldsymbol{x})\boldsymbol{u} = \sum_{i=1}^{n}\sum_{j=1}^{p} \left(\lambda_i \frac{\partial}{\partial x_1}\mathcal{B}_{ij}(\boldsymbol{x})\, u_j, \quad \ldots, \quad \lambda_i \frac{\partial}{\partial x_n}\mathcal{B}_{ij}(\boldsymbol{x})\, u_j \right). \tag{3.17}$$

The variation $\frac{\delta \bar{\mathcal{J}}}{\delta \boldsymbol{x}(t_1)} = \mathbf{0}$ with respect to the terminal state $\boldsymbol{x}(t_1)$ leads to the corresponding boundary condition for the co-state,

$$\boldsymbol{\lambda}^T(t_1) = \nabla M(\boldsymbol{x}(t_1), t_1) + \boldsymbol{\nu}^T \nabla \boldsymbol{\psi}(\boldsymbol{x}(t_1)). \tag{3.18}$$

Transposing finally gives

$$-\dot{\boldsymbol{\lambda}}(t) = \left(\nabla \boldsymbol{R}^T(\boldsymbol{x}(t)) + \boldsymbol{u}^T(t)\nabla \boldsymbol{\mathcal{B}}^T(\boldsymbol{x}(t))\right)\boldsymbol{\lambda}(t) + \left(\nabla L(\boldsymbol{x}(t), t)\right)^T, \tag{3.19}$$

$$\boldsymbol{\lambda}(t_1) = \nabla M^T(\boldsymbol{x}(t_1), t_1) + \nabla \boldsymbol{\psi}^T(\boldsymbol{x}(t_1))\,\boldsymbol{\nu}. \tag{3.20}$$

Note that the co-state $\boldsymbol{\lambda}(t)$ satisfies a terminal condition rather than an initial condition. The variation $\frac{\delta \bar{\mathcal{J}}}{\delta \boldsymbol{\nu}^T} = \mathbf{0}$ with respect to the Lagrange multipliers $\boldsymbol{\nu}^T$ finally yields

$$\boldsymbol{\psi}(\boldsymbol{x}(t_1)) = \mathbf{0}. \tag{3.21}$$

The state equation (3.11) and the adjoint equation (3.19) for the state $\boldsymbol{x}(t)$ and co-state $\boldsymbol{\lambda}(t)$, respectively, as well as the algebraic expression Eq. (3.10) for the vector of control signals $\boldsymbol{u}(t)$ constitute the *necessary optimality conditions.*

3.1.3. Optimal trajectory tracking

The optimal control problem considered in this thesis is to steer the system state $\boldsymbol{x}(t)$ as closely as possible along a desired reference trajectory $\boldsymbol{x}_d(t)$. A common choice of the cost functions L and M is the quadratic difference between the actual trajectory $\boldsymbol{x}(t)$ and desired trajectory $\boldsymbol{x}_d(t)$,

$$L(\boldsymbol{x}(t), t) = \frac{1}{2}(\boldsymbol{x}(t) - \boldsymbol{x}_d(t))^T \boldsymbol{\mathcal{S}}(\boldsymbol{x}(t) - \boldsymbol{x}_d(t)), \tag{3.22}$$

$$M(\boldsymbol{x}(t_1), t_1) = \frac{1}{2}(\boldsymbol{x}(t_1) - \boldsymbol{x}_1)^T \boldsymbol{\mathcal{S}}_1(\boldsymbol{x}(t_1) - \boldsymbol{x}_1). \tag{3.23}$$

The matrices $\boldsymbol{\mathcal{S}}$ and $\boldsymbol{\mathcal{S}}_1$ are usually assumed to be symmetric and positive definite matrices of weighting coefficients. The expression $\int_{t_0}^{t_1} dt \boldsymbol{x}^T(t)\boldsymbol{\mathcal{S}}\boldsymbol{x}(t)$ is called the weighted L^2 norm of $\boldsymbol{x}(t)$. The choice of Eqs. (3.22) and (3.23) for the cost functions defines the problem of *optimal trajectory tracking.*

Note that there are different possibilities for the terminal condition at time $t = t_1$. The notions end point and terminal point are used synonymously. A squared difference term $M(\boldsymbol{x}(t_1), t_1)$ as in Eq. (3.23) penalizes a large deviation of the terminal state $\boldsymbol{x}(t_1)$ from the state space point $\boldsymbol{x}_1$. An end point condition of the form $\boldsymbol{\psi}(\boldsymbol{x}(t_1)) = (\boldsymbol{x}(t_1) - \boldsymbol{x}_1)^T$ insists on $\boldsymbol{x}(t_1) = \boldsymbol{x}_1$ at the terminal time. Here, the latter case is called an *exact* or *sharp* terminal condition. Both possibilities can appear in the same problem. A sharp terminal condition is much more restrictive than a squared difference term. If $\boldsymbol{x}(t_1) = \boldsymbol{x}_1$ cannot be satisfied, a solution to the optimal control problem does not exist. Sharp terminal conditions require the controlled dynamical system to be controllable, i.e., there must exist at least one control signal enforcing a transfer from the initial state $\boldsymbol{x}(t_0) = \boldsymbol{x}_0$ to the terminal state $\boldsymbol{x}(t_1) = \boldsymbol{x}_1$. The case with no terminal conditions, $M(\boldsymbol{x}(t_1), t_1) \equiv 0$ and $\boldsymbol{\psi}(\boldsymbol{x}(t_1)) \equiv \mathbf{0}$, is called a *free* end point condition.

3.1.4. Discussion

The state equation (3.11) and the adjoint equation (3.19) for the state $\boldsymbol{x}(t)$ and co-state $\boldsymbol{\lambda}(t)$, respectively, as well as the expression Eq. (3.10) for the vector of control signals $\boldsymbol{u}(t)$ constitute the *necessary optimality conditions.* Note that the initial condition for the state equation is specified at the initial time t_0, while the initial condition for the adjoint equation is specified at the terminal time t_1. These mixed boundary conditions pose considerable difficulties for a numerical treatment. As will be discussed in Section 3.3, a straightforward numerical solution is not possible and one usually has to resort to an iterative scheme.

Even if it is possible to find a solution to the necessary optimality conditions, this solution can only be considered as a possible candidate solution for the problem of minimizing the target functional Eq. (3.1). Similar as for the problem of minimizing an ordinary function, the necessary optimality conditions only determine an extremum, and sufficient optimality conditions have to be employed to find out if this candidate indeed minimizes Eq. (3.1). However, the question of sufficiency is more subtle than for ordinary functions, see e.g. (Bryson and Ho, 1975) and (Liberzon, 2012). In this thesis, only necessary optimality conditions are discussed.

Here we discuss only the problem of minimizing a target functional Eq. (3.1) with a constraint in form of a controlled dynamical system. Other constraints in form of differential and algebraic equalities and inequalities can be introduced in optimal control. For technical applications, it is useful to introduce inequality constraints for control signals such that the control signal is not allowed to exceed or undershoot certain thresholds. For some problems, as e.g. unregularized optimal control problems, Eq. (3.1) with $\epsilon = 0$, the existence of a solution to the minimization problem can only be guaranteed if inequality constraints for the control signals are taken into account.

Other possible constraints are state constraints. For example, the dynamical sys-

tem might describe chemical reactions such that the state components are to be interpreted as the concentrations of some chemical species. Naturally, these concentrations must be positive quantities, and a control signal which decreases the value of a concentration below zero is physically impossible. While the controlled system alone might violate the condition of positivity, it is possible to enforce this condition in the context of optimal control.

Other variations of optimal control problems are sparse controls, which are especially useful for spatio-temporal control systems (Ryll, 2011; Casas and Tröltzsch, 2014). Sparse control means a term of the form $\int_{t_0}^{t_1} dt\, |\boldsymbol{u}(t)|$, with $|\boldsymbol{u}(t)| = \sqrt{\boldsymbol{u}^T(t)\,\boldsymbol{u}(t)}$, is added to the functional Eq. (3.1). This has the interesting effect that the control signal vanishes exactly for some time intervals, while it has a larger amplitude in others compared to a control without a sparsity term. In this way, it is possible to find out at which times the application of a control is most effective.

Finally, we comment on the role of the regularization term $\frac{\epsilon^2}{2}\int_{t_0}^{t_1} dt\,(\boldsymbol{u}(t))^2$, sometimes called a Tikhonov regularization, in the context of optimal trajectory tracking, i.e., for a target functional of the form

$$\mathcal{J}[\boldsymbol{x}(t), \boldsymbol{u}(t)] = \frac{1}{2}\int_{t_0}^{t_1} dt\,(\boldsymbol{x}(t) - \boldsymbol{x}_d(t))^T \boldsymbol{\mathcal{S}}\,(\boldsymbol{x}(t) - \boldsymbol{x}_d(t)) + \frac{1}{2}(\boldsymbol{x}(t_1) - \boldsymbol{x}_1)^T \boldsymbol{\mathcal{S}}_1\,(\boldsymbol{x}(t_1) - \boldsymbol{x}_1) + \frac{\epsilon^2}{2}\int_{t_0}^{t_1} dt\,(\boldsymbol{u}(t))^2 . \tag{3.24}$$

The general effect of the regularization term is to penalize large control values. Without any inequality constraints on the control, a finite value $\epsilon > 0$ is usually necessary to guarantee the existence of a solution to the minimization problem in terms of bounded and continuous state trajectories $\boldsymbol{x}(t)$. On the other hand, it usually increases the stability and accuracy of numerical computations of an optimal control. Nevertheless, the case $\epsilon = 0$, called an unregularized optimal control, is of special interest. For a fixed value of $\epsilon \geq 0$, among all *possible* control signals, the corresponding optimal control signal is the one which brings the controlled state closest to the desired trajectory $\boldsymbol{x}_d(t)$ as measured by Eq. (3.24). Furthermore, among all *optimal* controls, the optimal control for $\epsilon = 0$ brings the controlled state closest to the desired trajectory $\boldsymbol{x}_d(t)$. In other words: the distance measure Eq. (3.24), considered as a function of ϵ, has a minimum for $\epsilon = 0$. A proof of this fact is relatively simple. The total derivative of the augmented functional $\bar{\mathcal{J}}$ with respect

to ϵ is

$$\begin{aligned}
&\frac{d}{d\epsilon}\bar{\mathcal{J}}\left[\boldsymbol{x}\left(t\right),\boldsymbol{u}\left(t\right),\boldsymbol{\lambda}\left(t\right),\boldsymbol{\nu}\right]\\
=&\frac{\delta}{\delta\boldsymbol{x}\left(t\right)}\bar{\mathcal{J}}\left[\boldsymbol{x}\left(t\right),\boldsymbol{u}\left(t\right),\boldsymbol{\lambda}\left(t\right),\boldsymbol{\nu}\right]\frac{d}{d\epsilon}\boldsymbol{x}\left(t\right)+\frac{\delta}{\delta\boldsymbol{u}\left(t\right)}\bar{\mathcal{J}}\left[\boldsymbol{x}\left(t\right),\boldsymbol{u}\left(t\right),\boldsymbol{\lambda}\left(t\right),\boldsymbol{\nu}\right]\frac{d}{d\epsilon}\boldsymbol{u}\left(t\right)\\
&+\frac{\delta}{\delta\boldsymbol{\lambda}\left(t\right)}\bar{\mathcal{J}}\left[\boldsymbol{x}\left(t\right),\boldsymbol{u}\left(t\right),\boldsymbol{\lambda}\left(t\right),\boldsymbol{\nu}\right]\frac{d}{d\epsilon}\boldsymbol{\lambda}\left(t\right)+\frac{\delta}{\delta\boldsymbol{\nu}}\bar{\mathcal{J}}\left[\boldsymbol{x}\left(t\right),\boldsymbol{u}\left(t\right),\boldsymbol{\lambda}\left(t\right),\boldsymbol{\nu}\right]\frac{d}{d\epsilon}\boldsymbol{\nu}\\
&+\epsilon\int_{t_0}^{t_1}dt\left(\boldsymbol{u}\left(t\right)\right)^2. \qquad (3.25)
\end{aligned}$$

The last term $\frac{\partial}{\partial\epsilon}\bar{\mathcal{J}}=\epsilon\int_{t_0}^{t_1}dt\left(\boldsymbol{u}\left(t\right)\right)^2$ is due to the explicit dependence of $\bar{\mathcal{J}}$ on ϵ. If $\boldsymbol{\lambda}\left(t\right)$, $\boldsymbol{\lambda}\left(t_1\right)$, $\boldsymbol{u}\left(t\right)$, $\boldsymbol{\nu}$, and $\boldsymbol{x}\left(t\right)$ satisfy the necessary optimality conditions, Eq. (3.25) reduces to

$$\frac{d}{d\epsilon}\bar{\mathcal{J}}\left[\boldsymbol{x}\left(t\right),\boldsymbol{u}\left(t\right),\boldsymbol{\lambda}\left(t\right),\boldsymbol{\nu}\right]=\epsilon\int_{t_0}^{t_1}dt\left(\boldsymbol{u}\left(t\right)\right)^2. \qquad (3.26)$$

Then

$$\frac{d}{d\epsilon}\bar{\mathcal{J}}\left[\boldsymbol{x}\left(t\right),\boldsymbol{u}\left(t\right),\boldsymbol{\lambda}\left(t\right),\boldsymbol{\nu}\right]=0 \qquad (3.27)$$

for a non-vanishing control signal if and only if $\epsilon=0$. In other words, $\bar{\mathcal{J}}$ attains an extremum at $\epsilon=0$, and if $\bar{\mathcal{J}}$ is a minimum with respect to $\boldsymbol{\lambda}\left(t\right)$, $\boldsymbol{\lambda}\left(t_1\right)$, $\boldsymbol{u}\left(t\right)$, $\boldsymbol{\nu}$, and $\boldsymbol{x}\left(t\right)$, then it is also a minimum with respect to ϵ because of $\int_{t_0}^{t_1}dt\left(\boldsymbol{u}\left(t\right)\right)^2>0$.

Furthermore, any additional inequality constraints for state or control can only lead to a value of $\mathcal{J}$ smaller than or equal to its minimal value attained for $\epsilon=0$. The case with $\epsilon=0$ can be seen as the *limit of realizability* of a certain desired trajectory $\boldsymbol{x}_d\left(t\right)$. No other control, be it open or closed loop control, can enforce a state trajectory $\boldsymbol{x}\left(t\right)$ with a smaller distance to the desired state trajectory $\boldsymbol{x}_d\left(t\right)$ than an unregularized ($\epsilon=0$) optimal control. However, assuming $\epsilon=0$ leads to a singular optimal control problems involving additional difficulties.

3.2. Singular optimal control

Singular optimal control problems (Bell and Jacobson, 1975; Bryson and Ho, 1975) are best discussed in terms of the control Hamiltonian $H\left(\boldsymbol{x},\boldsymbol{u},t\right)$. As long as $H\left(\boldsymbol{x},\boldsymbol{u},t\right)$ depends only linearly on the vector of control signals $\boldsymbol{u}$, the optimal control problem is singular. For trajectory tracking tasks in affine control systems,

the control Hamiltonian is defined as

$$H(\boldsymbol{x},\boldsymbol{u},t) = \frac{1}{2}\left(\boldsymbol{x}(t)-\boldsymbol{x}_d(t)\right)^T \boldsymbol{\mathcal{S}}\left(\boldsymbol{x}(t)-\boldsymbol{x}_d(t)\right) + \frac{\epsilon^2}{2}|\boldsymbol{u}|^2 + \boldsymbol{\lambda}^T(t)\left(\boldsymbol{R}(\boldsymbol{x})+\boldsymbol{\mathcal{B}}(\boldsymbol{x})\boldsymbol{u}\right). \tag{3.28}$$

The control Hamiltonian is quadratic in the control signal $\boldsymbol{u}(t)$ as long as $\epsilon > 0$. The algebraic relation Eq. (3.29) between control signal and co-state is obtained from the condition of a stationary Hamiltonian with respect to control,

$$\boldsymbol{0} = \left(\nabla_{\boldsymbol{u}} H(\boldsymbol{x},\boldsymbol{u},t)\right)^T = \epsilon^2\boldsymbol{u} + \boldsymbol{\mathcal{B}}^T(\boldsymbol{x})\boldsymbol{\lambda}. \tag{3.29}$$

Clearly, if the regularization parameter $\epsilon = 0$, H depends only linearly on the control signal, and Eq. (3.29) reduces to

$$\boldsymbol{0} = \boldsymbol{\mathcal{B}}^T(\boldsymbol{x})\boldsymbol{\lambda}. \tag{3.30}$$

While Eq. (3.29) can be used to obtain the control signal $\boldsymbol{u}(t)$ in terms of the state $\boldsymbol{x}(t)$ and co-state $\boldsymbol{\lambda}(t)$, this is clearly impossible for Eq. (3.30). Additional necessary optimality condition, known as the Kelly or generalized Legendre-Clebsch condition, must be employed to determine an expression for the control signal.

3.2.1. The Kelly condition

It can be rigorously proven that additional necessary optimality condition besides the usual necessary optimality conditions have to be satisfied in case of singular optimal controls (Bell and Jacobson, 1975). This condition is known as the Kelly condition in case of single-component control signals and as the generalized Legendre-Clebsch condition in case of multi-component control signals. For simplicity, only the case of scalar control signals $u(t)$ is considered in this section. The singular control Hamiltonian for optimal trajectory tracking, Eq. (3.28) with $\epsilon = 0$, becomes

$$H(\boldsymbol{x},u,t) = \frac{1}{2}\left(\boldsymbol{x}(t)-\boldsymbol{x}_d(t)\right)^T \boldsymbol{\mathcal{S}}\left(\boldsymbol{x}(t)-\boldsymbol{x}_d(t)\right) + \boldsymbol{\lambda}^T\left(\boldsymbol{R}(\boldsymbol{x})+\boldsymbol{B}(\boldsymbol{x})u\right). \tag{3.31}$$

The Kelly condition is (Bell and Jacobson, 1975)

$$(-1)^k \frac{\partial}{\partial u}\left[\frac{d^{2k}}{dt^{2k}}\frac{\partial}{\partial u}H(\boldsymbol{x}(t),u(t),t)\right] \geq 0,\ k = 1,2,\ldots, \tag{3.32}$$

and is utilized as follows. The stationarity condition $\partial_u H = 0$, or, equivalently,

$$0 = \boldsymbol{\lambda}^T(t)\boldsymbol{B}(\boldsymbol{x}(t)), \tag{3.33}$$

is valid for all times t but cannot be used to obtain an expression for the control signal $u(t)$. Applying the time derivative to Eq. (3.33) yields $\frac{d}{dt}\partial_u H = 0$, or

$$\begin{aligned} 0 &= \dot{\boldsymbol{\lambda}}^T(t)\,\boldsymbol{B}(\boldsymbol{x}(t)) + \boldsymbol{\lambda}^T(t)\,\nabla\boldsymbol{B}(\boldsymbol{x}(t))\,\dot{\boldsymbol{x}}(t) \\ &= \boldsymbol{\lambda}^T(t)\,\boldsymbol{q}(\boldsymbol{x}(t)) - (\boldsymbol{x}(t) - \boldsymbol{x}_d(t))^T\,\boldsymbol{\mathcal{S}}\boldsymbol{B}(\boldsymbol{x}(t)). \end{aligned} \tag{3.34}$$

The controlled state equation as well as the co-state equations was used and $\boldsymbol{q}(\boldsymbol{x})$ denotes the abbreviation

$$\boldsymbol{q}(\boldsymbol{x}) = \nabla\boldsymbol{B}(\boldsymbol{x})\,\boldsymbol{R}(\boldsymbol{x}) - \nabla\boldsymbol{R}(\boldsymbol{x})\,\boldsymbol{B}(\boldsymbol{x}). \tag{3.35}$$

Equation (3.34) yields an additional relation between state $\boldsymbol{x}$ and co-state $\boldsymbol{\lambda}$, but does not depend on the control signal $u(t)$. Therefore, it cannot be used to obtain an expression for $u(t)$. Applying the second time derivative $\frac{d^2}{dt^2}\partial_u H = 0$ to the stationarity condition yields

$$\begin{aligned} \frac{d^2}{dt^2}\partial_u H &= \dot{\boldsymbol{\lambda}}^T\boldsymbol{q}(\boldsymbol{x}) + \boldsymbol{\lambda}^T\nabla\boldsymbol{q}(\boldsymbol{x})\,\dot{\boldsymbol{x}} - (\dot{\boldsymbol{x}} - \dot{\boldsymbol{x}}_d)^T\,\boldsymbol{\mathcal{S}}\boldsymbol{B}(\boldsymbol{x}) - (\boldsymbol{x} - \boldsymbol{x}_d)^T\,\boldsymbol{\mathcal{S}}\nabla\boldsymbol{B}(\boldsymbol{x})\,\dot{\boldsymbol{x}} \\ &= \boldsymbol{\lambda}^T(\nabla\boldsymbol{q}(\boldsymbol{x})\,\boldsymbol{R}(\boldsymbol{x}) - \nabla\boldsymbol{R}(\boldsymbol{x})\,\boldsymbol{q}(\boldsymbol{x})) + \boldsymbol{B}^T(\boldsymbol{x})\,\boldsymbol{\mathcal{S}}(\dot{\boldsymbol{x}}_d - \boldsymbol{R}(\boldsymbol{x})) \\ &\quad - (\boldsymbol{x} - \boldsymbol{x}_d)^T\,\boldsymbol{\mathcal{S}}(\nabla\boldsymbol{B}(\boldsymbol{x})\,\boldsymbol{R}(\boldsymbol{x}) + \boldsymbol{q}(\boldsymbol{x})) + p(\boldsymbol{x})\,u. \end{aligned} \tag{3.36}$$

Here, $p(\boldsymbol{x})$ denotes the abbreviation

$$\begin{aligned} p(\boldsymbol{x}) &= \boldsymbol{\lambda}^T(\nabla\boldsymbol{q}(\boldsymbol{x})\,\boldsymbol{B}(\boldsymbol{x}) - \nabla\boldsymbol{B}(\boldsymbol{x})\,\boldsymbol{q}(\boldsymbol{x})) - \boldsymbol{B}^T(\boldsymbol{x})\,\boldsymbol{\mathcal{S}}\boldsymbol{B}(\boldsymbol{x}) \\ &\quad - (\boldsymbol{x} - \boldsymbol{x}_d)^T\,\boldsymbol{\mathcal{S}}\nabla\boldsymbol{B}(\boldsymbol{x})\,\boldsymbol{B}(\boldsymbol{x}). \end{aligned} \tag{3.37}$$

Equation (3.36) does depend on the control, and as long as $p(\boldsymbol{x}) \neq 0$, it can be solved for the control signal u,

$$\begin{aligned} u &= \frac{1}{p(\boldsymbol{x})}\left(\boldsymbol{\lambda}^T(\nabla\boldsymbol{q}(\boldsymbol{x})\,\boldsymbol{R}(\boldsymbol{x}) - \nabla\boldsymbol{R}(\boldsymbol{x})\,\boldsymbol{q}(\boldsymbol{x})) + \boldsymbol{B}^T(\boldsymbol{x})\,\boldsymbol{\mathcal{S}}(\dot{\boldsymbol{x}}_d - \boldsymbol{R}(\boldsymbol{x}))\right) \\ &\quad - \frac{1}{p(\boldsymbol{x})}(\boldsymbol{x} - \boldsymbol{x}_d)^T\,\boldsymbol{\mathcal{S}}(\nabla\boldsymbol{B}(\boldsymbol{x})\,\boldsymbol{R}(\boldsymbol{x}) + \boldsymbol{q}(\boldsymbol{x})). \end{aligned} \tag{3.38}$$

If $\frac{d^2}{dt^2}\partial_u H$ would not depend on u, the time derivative must be applied repeatedly to the stationarity condition $\partial_u H$ until an expression depending on u is generated. It can be shown that scalar control signals only appear in even orders of the total time derivative. This is the reason for the term $\frac{d^{2k}}{dt^{2k}}$ in the Kelly conditions Eq. (3.32) (Bell and Jacobson, 1975). Finally, one has to check for a generalized convexity condition,

$$(-1)^k\,\partial_u\left[\frac{d^{2k}}{dt^{2k}}\partial_u H\right] > 0. \tag{3.39}$$

For Eq. (3.36) with $k = 1$, the generalized convexity condition

$$-\partial_u \left[\frac{d^2}{dt^2} \partial_u H \right] = p(\boldsymbol{x}) > 0. \tag{3.40}$$

Thus, as long as $p(\boldsymbol{x}) > 0$, the control signal given by Eq. (3.38) satisfies all necessary optimality conditions. The procedure is discussed with a simple example, namely a mechanical control system in one spatial dimension with vanishing external force.

Example 3.1: Singular optimal control of a free particle

Consider the Newton's equation of motion for the position x of a free point mass in one spatial dimension under the influence of a control force u,

$$\ddot{x}(t) = u(t). \tag{3.41}$$

Written as a dynamical system, Eq. (3.41) becomes

$$\dot{x}(t) = y(t), \tag{3.42}$$
$$\dot{y}(t) = u(t). \tag{3.43}$$

The optimal control task is to minimize the constrained functional Eq. (3.24) without any regularization term. For simplicity, the desired trajectories are chosen to vanish,

$$x_d(t) \equiv 0, \qquad y_d(t) \equiv 0. \tag{3.44}$$

The initial time is set to $t_0 = 0$. The optimal control task reduces to the minimization of the functional

$$\mathcal{J}[\boldsymbol{x}(t), u(t)] = \frac{1}{2} \int_0^{t_1} \left((x(t))^2 + (y(t))^2 \right) dt, \tag{3.45}$$

subject to the dynamics Eqs. (3.42) and (3.43) with sharp terminal conditions and zero initial conditions,

$$x(0) = 0, \qquad y(0) = 0, \tag{3.46}$$
$$x(t_1) = x_1, \qquad y(t_1) = y_1. \tag{3.47}$$

Thus, the state is required to reach the terminal state $\boldsymbol{x}_1 = \begin{pmatrix} x_1, & y_1 \end{pmatrix}^T$ exactly.

The necessary optimality conditions are

$$\begin{pmatrix} \dot{x}(t) \\ \dot{y}(t) \end{pmatrix} = \begin{pmatrix} y(t) \\ 0 \end{pmatrix} + \begin{pmatrix} 0 \\ u(t) \end{pmatrix}, \tag{3.48}$$

$$-\begin{pmatrix} \dot{\lambda}_x(t) \\ \dot{\lambda}_y(t) \end{pmatrix} = \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} \lambda_x(t) \\ \lambda_y(t) \end{pmatrix} + \begin{pmatrix} x(t) \\ y(t) \end{pmatrix}. \tag{3.49}$$

The stationarity condition $\partial_u H = 0$ becomes

$$\lambda_y(t) = 0. \tag{3.50}$$

Applying the time derivative to Eq. (3.50), and using the co-state equation to eliminate $\dot{\lambda}_y(t)$ yields a relation between λ_x and y as

$$0 = \frac{d}{dt}\partial_u H = \dot{\lambda}_y(t) = -\lambda_x(t) - y(t). \tag{3.51}$$

Applying the second order time derivative to the stationary condition Eq. (3.50),

$$0 = \frac{d^2}{dt^2}\partial_u H = -\dot{\lambda}_x(t) - \dot{y}(t) = x(t) - u(t), \tag{3.52}$$

yields an expression for the control signal

$$u(t) = x(t). \tag{3.53}$$

Finally, the generalized convexity condition yields

$$-\partial_u \left[\frac{d^2}{dt^2}\partial_u H\right] = 1 > 0, \tag{3.54}$$

and all conditions encoded in the Kelly condition are satisfied. In summary, the co-states are governed by two algebraic equations

$$\lambda_x(t) = -y(t), \qquad \lambda_y(t) = 0, \tag{3.55}$$

while the control signal is obtained in terms of the controlled state component x as

$$u(t) = x(t). \tag{3.56}$$

The state components x and y are governed by two coupled linear differential equations of first order,

$$\dot{x}(t) = y(t), \qquad \dot{y}(t) = x(t). \tag{3.57}$$

The general solution to Eqs. (3.57), with two constants of integration C_1 and C_2, is

$$x(t) = C_1 \cosh(t) + C_2 \sinh(t), \tag{3.58}$$
$$y(t) = C_1 \sinh(t) + C_2 \cosh(t). \tag{3.59}$$

However, only two out of four boundary conditions Eqs. (3.46) and (3.47) can be satisfied the solution. How to resolve this problem is demonstrated later on by investigating the same problem in the limit of small regularization parameter $\epsilon \to 0$.

3.3. Numerical solution of optimal control problems

A numerical solution of an optimal control problem requires the simultaneous solution of

1. the n state variables $\boldsymbol{x}(t)$ as solution to the controlled state equation (3.11),
2. the n adjoint state variables $\boldsymbol{\lambda}(t)$ as solution to the adjoint equation (3.19),
3. the p algebraic equations for the control vector $\boldsymbol{u}(t)$ as given by (3.10).

A straightforward solution of this coupled system of equations is rarely possible. The problem is that the boundary conditions for the controlled state equation, $\boldsymbol{x}(t_0) = \boldsymbol{x}_0$, are given at the initial time $t = t_0$, while the adjoint equation is to be solved with the terminal condition $\boldsymbol{\lambda}(t_1) = \nabla M^T(\boldsymbol{x}(t_1)) + \nabla \boldsymbol{\psi}^T(\boldsymbol{x}(t_1))\boldsymbol{\nu}$, Eq. (3.20), given at the terminal time $t = t_1$. This typically requires an iterative solution algorithm similar to the shooting method.

Another problem is that the adjoint equation (3.19) yields an unstable time evolution, which in turn leads to an unstable numerical algorithm. This problem can be tackled by solving the time-reversed adjoint equation. Introducing the new time

$$\tilde{t} = t_1 - t, \qquad 0 \le \tilde{t} \le t_1 - t_0, \tag{3.60}$$

and new adjoint state variables

$$\tilde{\boldsymbol{\lambda}}\left(\tilde{t}\right) = \tilde{\boldsymbol{\lambda}}(t_1 - t) = \boldsymbol{\lambda}(t), \tag{3.61}$$

the adjoint equation (3.19) is transformed to a new equation

$$\begin{aligned}\dot{\tilde{\boldsymbol{\lambda}}}\left(\tilde{t}\right) = &\left(\nabla \boldsymbol{R}^T\left(\boldsymbol{x}\left(t_1 - \tilde{t}\right)\right) + \boldsymbol{u}^T\left(t_1 - \tilde{t}\right)\nabla \boldsymbol{\mathcal{B}}^T\left(\boldsymbol{x}\left(t_1 - \tilde{t}\right)\right)\right)\tilde{\boldsymbol{\lambda}}\left(\tilde{t}\right) \\ &+ \left(\nabla L\left(\boldsymbol{x}\left(t_1 - \tilde{t}\right), t_1 - \tilde{t}\right)\right)^T,\end{aligned} \tag{3.62}$$

with initial condition

$$\tilde{\boldsymbol{\lambda}}(0) = \nabla M^T(\boldsymbol{x}(t_1)) + \nabla \boldsymbol{\psi}^T(\boldsymbol{x}(t_1))\boldsymbol{\nu}. \tag{3.63}$$

If the original equation (3.19) yields an unstable time evolution, then Eq. (3.62) yields a stable time evolution, and vice versa. Although the transformed adjoint equation (3.62) as well as the controlled state equation (3.11) are both to be solved with initial conditions, they cannot be solved straightforwardly: while $\tilde{\boldsymbol{\lambda}}$ in Eq. (3.62) is evaluated at $\tilde{t}$, the state variable $\boldsymbol{x}$ is evaluated at the reversed time $t_1 - \tilde{t}$. This again illustrates the problem posed by boundary conditions defined at different times. These difficulties cannot be resolved by a simple time reversion of the adjoint equation.

A relatively simple algorithm to solve the optimal control problem is a first order gradient algorithm. Starting with an initial guess for the control as e.g. $\boldsymbol{u}(t) \equiv \boldsymbol{0}$, the iterative algorithm proceeds as follows (the integer k denotes the k-th iterate):

1. solve the controlled state equation to obtain the controlled state $\boldsymbol{x}^k(t)$ from Eq. (3.11),
2. use $\boldsymbol{x}^k(t)$ in Eq. (3.19) to obtain $\boldsymbol{\lambda}^k(t)$
3. compute a new control $\boldsymbol{u}^{k+1}(t)$ with the help of Eq. (3.10) as

$$\boldsymbol{u}^{k+1}(t) = \boldsymbol{u}^k(t) - s\left(\epsilon^2 \boldsymbol{u}^k(t) + \boldsymbol{\mathcal{B}}^T\left(\boldsymbol{x}^k(t)\right)\boldsymbol{\lambda}^k(t)\right) \tag{3.64}$$

4. set $k = k + 1$ and go to 1.

The idea of this iterative algorithm is to change the control in the correct "direction" in function space such that the control converges to the optimal solution. The step width s is an important quantity. It is often chosen adaptively, with large step widths for the first couple of iterations and progressively smaller step widths as the solution for the control converges. Depending on the type of problem, some hundred up to many hundred thousand of iterations have to be performed to find a sufficiently correct solution. This renders optimal control algorithms computationally expensive, and prevents application of optimizations in real time for processes which are too fast.

The ACADO Toolkit (Houska et al., 2013, 2011a,b) is a readily available open source package to solve optimal control problems. If not stated otherwise, this toolkit is used for all numerical solutions of optimal control throughout the thesis. A typical computation for a dynamical system with two state components on a time interval of length 1 and step width $\Delta t = 10^{-3}$, as shown in Example 1.4, takes about half an hour on a standard laptop.

Many varieties and improvements of the algorithm sketched above can be found in the literature, as e.g. conjugated gradient method, see (Shewchuk, 1994) and references therein. Other algorithms to solve optimal control problems exist, as e.g. the Newton-Raphson root finding algorithm, see (Nocedal and Wright, 2006) and (Bryson and Ho, 1975) for an overview and examples.

3.4. Exactly realizable trajectories and optimal control

This section discusses the conditions under which exactly realizable trajectories are optimal. In particular, an exactly realizable desired trajectory together with its corresponding control signal satisfies all necessary optimality conditions of a singular optimal control problem.

3.4.1. Usual necessary optimality conditions

The approach to control in terms of exactly realizable trajectories from Chapter 2 is closely related to an unregularized optimal control problem. Let the target functional be

$$\begin{aligned}\mathcal{J}\left[\boldsymbol{x}\left(t\right),\boldsymbol{u}\left(t\right)\right] =& \frac{1}{2}\int_{t_0}^{t_1} dt\left(\boldsymbol{x}\left(t\right)-\boldsymbol{x}_d\left(t\right)\right)^T\boldsymbol{\mathcal{S}}\left(\boldsymbol{x}\left(t\right)-\boldsymbol{x}_d\left(t\right)\right)\\ &+\frac{1}{2}\left(\boldsymbol{x}\left(t_1\right)-\boldsymbol{x}_1\right)\boldsymbol{\mathcal{S}}_1\left(\boldsymbol{x}\left(t_1\right)-\boldsymbol{x}_1\right)+\frac{\epsilon^2}{2}\int_{t_0}^{t_1} dt\left|\boldsymbol{u}\left(t\right)\right|^2. \end{aligned} \tag{3.65}$$

Here, $\boldsymbol{x}_d\left(t\right)$ is the desired trajectory, $\boldsymbol{\mathcal{S}}$ and $\boldsymbol{\mathcal{S}}_1$ are symmetric positive definite matrices, and $\boldsymbol{x}_1$ is a desired terminal state. The necessary optimality conditions comprise the controlled state equation

$$\dot{\boldsymbol{x}}\left(t\right)=\boldsymbol{R}\left(\boldsymbol{x}\left(t\right)\right)+\boldsymbol{\mathcal{B}}\left(\boldsymbol{x}\left(t\right)\right)\boldsymbol{u}\left(t\right), \tag{3.66}$$

$$\boldsymbol{x}\left(t_0\right)=\boldsymbol{x}_0, \tag{3.67}$$

the adjoint equation

$$-\dot{\boldsymbol{\lambda}}\left(t\right)=\left(\nabla\boldsymbol{R}^T\left(\boldsymbol{x}\left(t\right)\right)+\boldsymbol{u}^T\left(t\right)\nabla\boldsymbol{\mathcal{B}}^T\left(\boldsymbol{x}\left(t\right)\right)\right)\boldsymbol{\lambda}\left(t\right)+\boldsymbol{\mathcal{S}}\left(\boldsymbol{x}\left(t\right)-\boldsymbol{x}_d\left(t\right)\right), \tag{3.68}$$

$$\boldsymbol{\lambda}\left(t_1\right)=\boldsymbol{\mathcal{S}}_1\left(\boldsymbol{x}\left(t_1\right)-\boldsymbol{x}_1\right), \tag{3.69}$$

and an algebraic relation between co-state $\boldsymbol{\lambda}$ and control $\boldsymbol{u}$,

$$\epsilon^2\boldsymbol{u}\left(t\right)+\boldsymbol{\mathcal{B}}^T\left(\boldsymbol{x}\left(t\right)\right)\boldsymbol{\lambda}\left(t\right)=0. \tag{3.70}$$

Usually, Eq. (3.70) is used to determine the control $\boldsymbol{u}\left(t\right)$. Here, we proceed differently, and assume an exactly realizable desired trajectory $\boldsymbol{x}_d\left(t\right)$ such that the controlled state $\boldsymbol{x}\left(t\right)$ exactly follows $\boldsymbol{x}_d\left(t\right)$ for all times,

$$\boldsymbol{x}\left(t\right)=\boldsymbol{x}_d\left(t\right). \tag{3.71}$$

Starting from this assumption, the necessary optimality conditions are evaluated to determine conditions on $\boldsymbol{x}_d\left(t\right)$ such that Eq. (3.71) holds.

First of all, for Eq. (3.71) to be valid at all times, the initial value of the desired trajectory must comply with the initial state,

$$\boldsymbol{x}\left(t_0\right) = \boldsymbol{x}_0 = \boldsymbol{x}_d\left(t_0\right). \tag{3.72}$$

With assumption Eq. (3.71), the adjoint equation becomes

$$-\dot{\boldsymbol{\lambda}}\left(t\right) = \left(\nabla \boldsymbol{R}^T\left(\boldsymbol{x}_d\left(t\right)\right) + \boldsymbol{u}^T\left(t\right)\nabla \boldsymbol{\mathcal{B}}^T\left(\boldsymbol{x}_d\left(t\right)\right)\right)\boldsymbol{\lambda}\left(t\right), \tag{3.73}$$

$$\boldsymbol{\lambda}\left(t_1\right) = \boldsymbol{\mathcal{S}}_1\left(\boldsymbol{x}_d\left(t_1\right) - \boldsymbol{x}_1\right). \tag{3.74}$$

If the desired trajectory satisfies

$$\boldsymbol{x}_d\left(t_1\right) = \boldsymbol{x}_1, \tag{3.75}$$

the boundary condition for the adjoint equation becomes

$$\boldsymbol{\lambda}\left(t_1\right) = \boldsymbol{0}. \tag{3.76}$$

Consequently, the co-state $\boldsymbol{\lambda}$ vanishes identically for all times,

$$\boldsymbol{\lambda}\left(t\right) \equiv \boldsymbol{0}. \tag{3.77}$$

With the help of the Moore-Penrose pseudo inverse

$$\boldsymbol{\mathcal{B}}^{+}\left(\boldsymbol{x}\right) = \left(\boldsymbol{\mathcal{B}}^T\left(\boldsymbol{x}\right)\boldsymbol{\mathcal{B}}\left(\boldsymbol{x}\right)\right)^{-1}\boldsymbol{\mathcal{B}}^T\left(\boldsymbol{x}\right), \tag{3.78}$$

the controlled state equation (3.66) is solved for the control signal

$$\boldsymbol{u}\left(t\right) = \boldsymbol{\mathcal{B}}^{+}\left(\boldsymbol{x}_d\left(t\right)\right)\left(\dot{\boldsymbol{x}}_d\left(t\right) - \boldsymbol{R}\left(\boldsymbol{x}_d\left(t\right)\right)\right). \tag{3.79}$$

Using $\boldsymbol{u}\left(t\right)$ in the controlled state equation yields the constraint equation

$$\boldsymbol{0} = \boldsymbol{\mathcal{Q}}\left(\boldsymbol{x}_d\left(t\right)\right)\left(\dot{\boldsymbol{x}}_d\left(t\right) - \boldsymbol{R}\left(\boldsymbol{x}_d\left(t\right)\right)\right). \tag{3.80}$$

Finally, because of the vanishing co-state Eq. (3.77), the stationarity condition Eq. (3.70) becomes

$$\epsilon^2\boldsymbol{u}\left(t\right) = 0. \tag{3.81}$$

Clearly, because $\boldsymbol{u}\left(t\right)$ is non-vanishing, Eq. (3.81) can only be satisfied if

$$\epsilon = 0. \tag{3.82}$$

In conclusion, for the necessary optimality conditions to be valid under the assumption $\boldsymbol{x}\left(t\right) = \boldsymbol{x}_d\left(t\right)$, the desired trajectory $\boldsymbol{x}_d\left(t\right)$ has to be exactly realizable. Furthermore, $\boldsymbol{x}_d\left(t\right)$ must comply with the terminal condition $\boldsymbol{x}_d\left(t_1\right) = \boldsymbol{x}_1$. This additional condition originates simply from different formulations of the control task. While the control methods from Chapter 2 enforce only initial conditions, optimal control is able to impose terminal conditions as well. As shown by Eq. (3.82), exactly realizable trajectories naturally lead to singular optimal control problems. Hence, additional necessary optimality conditions in form of the generalized Legendre-Clebsch conditions must be evaluated.

3.4.2. The generalized Legendre-Clebsch conditions

The generalized Legendre-Clebsch conditions are (Bell and Jacobson, 1975)

$$\nabla_{\boldsymbol{u}} \frac{d^k}{dt^k} \left(\nabla_{\boldsymbol{u}} H\right) = \mathbf{0},\ k \in \mathbb{N},\ k \text{ odd}, \tag{3.83}$$

and

$$(-1)^l \nabla_{\boldsymbol{u}} \frac{d^{2l}}{dt^{2l}} \left(\nabla_{\boldsymbol{u}} H\right) \geq \mathbf{0},\ l \in \mathbb{N}. \tag{3.84}$$

These conditions are evaluated in the same manner as the Kelly condition, see Section 3.2.1. Due to the vector character of the control signal, the computations are more involved, and a different notation is adopted. Written for the individual state components x_i, the controlled state equation is

$$\dot{x}_i = R_i + \sum_{k=1}^{p} \mathcal{B}_{ik} u_k. \tag{3.85}$$

For the remainder of this section, the state arguments of $\boldsymbol{R}$ and $\boldsymbol{\mathcal{B}}$ and time arguments of $\boldsymbol{\lambda}$, $\boldsymbol{x}$, $\boldsymbol{x}_d$, and $\boldsymbol{u}$ are suppressed to shorten the notation. The matrix entries $\mathcal{B}_{ik}$ are assumed to depend on the state $\boldsymbol{x}$. With

$$\partial_j = \frac{\partial}{\partial x_j}, \qquad \left(\nabla \boldsymbol{R}\right)_{ij} = \partial_j R_i, \qquad \left(\boldsymbol{\lambda}^T \nabla \boldsymbol{\mathcal{B}} \boldsymbol{u}\right)_k = \sum_{i=1}^{n} \sum_{j=1}^{p} \lambda_i \partial_k \mathcal{B}_{ij} u_j, \tag{3.86}$$

the adjoint equation is written as

$$-\dot{\lambda}_i = \sum_{k=1}^{n} \left(\lambda_k \partial_i R_k + \sum_{l=1}^{p} \lambda_k \partial_i \mathcal{B}_{kl} u_l + \left(x_k - x_{d,k}\right) \mathcal{S}_{ki} \right). \tag{3.87}$$

Here, $x_{d,k}$ denotes the k-th component of the desired trajectory $\boldsymbol{x}_d\left(t\right)$. The stationarity condition

$$\mathbf{0} = \nabla_{\boldsymbol{u}} H = \boldsymbol{\lambda}^T \boldsymbol{\mathcal{B}} \tag{3.88}$$

becomes

$$0 = \sum_{i=1}^{n} \lambda_i \mathcal{B}_{ij}. \tag{3.89}$$

The procedure is analogous to the Kelly condition and the time derivative is applied repeatedly onto Eq. (3.89). The first condition, Eq. (3.83) for $k = 1$,

$$\frac{d}{dt} \left(\nabla_{\boldsymbol{u}} H\right) = \mathbf{0}, \tag{3.90}$$

yields

$$\begin{aligned}
0 &= \sum_{i=1}^{n} \left(\dot{\lambda}_i \mathcal{B}_{ij} + \sum_{k=1}^{n} \lambda_i \partial_k \mathcal{B}_{ij} \dot{x}_k \right) \\
&= \sum_{l=1}^{p} \sum_{i=1}^{n} \sum_{k=1}^{n} \lambda_i \left(\partial_k \mathcal{B}_{ij} \mathcal{B}_{kl} - \mathcal{B}_{kj} \partial_k \mathcal{B}_{il} \right) u_l - \sum_{i=1}^{n} \sum_{k=1}^{n} \left(x_k - x_{d,k} \right) \mathcal{S}_{ki} \mathcal{B}_{ij} \\
&+ \sum_{i=1}^{n} \sum_{k=1}^{n} \lambda_i \left(\partial_k \mathcal{B}_{ij} R_k - \partial_k R_i \mathcal{B}_{kj} \right). \qquad (3.91)
\end{aligned}$$

Because of $\boldsymbol{\lambda}(t) \equiv \mathbf{0}$ and $\boldsymbol{x}(t) = \boldsymbol{x}_d(t)$ for all times, this expression is satisfied. Note that the condition

$$\nabla_{\boldsymbol{u}} \frac{d}{dt} \left(\nabla_{\boldsymbol{u}} H \right) = \mathbf{0}, \qquad (3.92)$$

or

$$0 = \sum_{i=1}^{n} \sum_{k=1}^{n} \lambda_i \left(\partial_k \mathcal{B}_{ij} \mathcal{B}_{kl} - \mathcal{B}_{kj} \partial_k \mathcal{B}_{il} \right), \qquad (3.93)$$

is only valid under certain symmetry conditions on the coupling matrix $\boldsymbol{\mathcal{B}}(\boldsymbol{x})$ for a finite co-state $\lambda_i \neq 0$ (Bryson and Ho, 1975). However, here the co-state vanishes exactly, and it is unnecessary to impose these symmetry conditions on $\boldsymbol{\mathcal{B}}$. The next Legendre-Clebsch condition is

$$\frac{d^2}{dt^2} \left(\nabla_{\boldsymbol{u}} H \right) = \mathbf{0}, \qquad (3.94)$$

or

$$\begin{aligned}
0 &= \sum_{i=1}^{n} \sum_{k=1}^{n} \left(\sum_{l=1}^{p} \dot{\lambda}_i \left(\partial_k \mathcal{B}_{ij} \mathcal{B}_{kl} - \mathcal{B}_{kj} \partial_k \mathcal{B}_{il} \right) u_l + \sum_{l=1}^{p} \lambda_i \frac{d}{dt} \left(\partial_k \mathcal{B}_{ij} \mathcal{B}_{kl} - \mathcal{B}_{kj} \partial_k \mathcal{B}_{il} \right) u_l \right) \\
&+ \sum_{i=1}^{n} \sum_{k=1}^{n} \sum_{l=1}^{p} \lambda_i \left(\partial_k \mathcal{B}_{ij} \mathcal{B}_{kl} - \mathcal{B}_{kj} \partial_k \mathcal{B}_{il} \right) \dot{u}_l \\
&- \sum_{i=1}^{n} \sum_{k=1}^{n} \left(\left(R_k + \sum_{m=1}^{p} \mathcal{B}_{km} u_m - \dot{x}_{d,k} \right) \mathcal{S}_{ki} \mathcal{B}_{ij} + \sum_{m=1}^{n} \left(x_k - x_{d,k} \right) \mathcal{S}_{ki} \partial_m \mathcal{B}_{ij} \dot{x}_m \right) \\
&+ \sum_{i=1}^{n} \sum_{k=1}^{n} \left(\dot{\lambda}_i \left(\partial_k \mathcal{B}_{ij} R_k - \partial_k R_i \mathcal{B}_{kj} \right) + \lambda_i \frac{d}{dt} \left(\partial_k \mathcal{B}_{ij} R_k - \partial_k R_i \mathcal{B}_{kj} \right) \right). \qquad (3.95)
\end{aligned}$$

The controlled state equation (3.85) was used to substitute $\dot{x}_k$. Because of $\boldsymbol{\lambda}(t) \equiv \mathbf{0}$ and $\boldsymbol{x}(t) = \boldsymbol{x}_d(t)$ for all times, the Eq. (3.95) simplifies considerably and yields a solution for the control signal

$$\sum_{i=1}^{n} \sum_{k=1}^{n} \sum_{m=1}^{p} \mathcal{B}_{ij} \mathcal{S}_{ki} \mathcal{B}_{km} u_m = \sum_{i=1}^{n} \sum_{k=1}^{n} \mathcal{B}_{ij} \mathcal{S}_{ki} \left(\dot{x}_{d,k} - R_k \right). \qquad (3.96)$$

Casting Eq. (3.96) in terms of vectors and matrices and exploiting the symmetry of $\boldsymbol{\mathcal{S}}$ gives

$$\boldsymbol{\mathcal{B}}^T \boldsymbol{\mathcal{S}} \boldsymbol{\mathcal{B}} \boldsymbol{u} = \boldsymbol{\mathcal{B}}^T \boldsymbol{\mathcal{S}} \left(\dot{\boldsymbol{x}}_d - \boldsymbol{R} \right). \tag{3.97}$$

Solving for the control and substituting $\boldsymbol{x}(t) = \boldsymbol{x}_d(t)$ results in

$$\boldsymbol{u}(t) = \boldsymbol{\mathcal{B}}^g_{\boldsymbol{\mathcal{S}}} \left(\boldsymbol{x}_d(t) \right) \left(\dot{\boldsymbol{x}}_d(t) - \boldsymbol{R} \left(\boldsymbol{x}_d(t) \right) \right). \tag{3.98}$$

The $p \times n$ matrix $\boldsymbol{\mathcal{B}}^g_{\boldsymbol{\mathcal{S}}}(\boldsymbol{x})$ is defined by

$$\boldsymbol{\mathcal{B}}^g_{\boldsymbol{\mathcal{S}}}(\boldsymbol{x}) = \left(\boldsymbol{\mathcal{B}}^T(\boldsymbol{x}) \boldsymbol{\mathcal{S}} \boldsymbol{\mathcal{B}}(\boldsymbol{x}) \right)^{-1} \boldsymbol{\mathcal{B}}^T(\boldsymbol{x}) \boldsymbol{\mathcal{S}}. \tag{3.99}$$

Note that the matrix $\boldsymbol{\mathcal{B}}^g_{\boldsymbol{\mathcal{S}}}(\boldsymbol{x})$ is not the Moore-Penrose pseudo inverse but a generalized reflexive inverse of $\boldsymbol{\mathcal{B}}(\boldsymbol{x})$, see Appendix A.2.1. Finally, the generalized convexity condition

$$\nabla_{\boldsymbol{u}} \frac{d^2}{dt^2} \left(\nabla_{\boldsymbol{u}} H \right) \geq \boldsymbol{0} \tag{3.100}$$

is satisfied whenever

$$\boldsymbol{\mathcal{B}}^T(\boldsymbol{x}) \boldsymbol{\mathcal{S}} \boldsymbol{\mathcal{B}}(\boldsymbol{x}) > \boldsymbol{0} \tag{3.101}$$

for all $\boldsymbol{x}$. Condition Eq. (3.101) ensures that $\left(\boldsymbol{\mathcal{B}}^T(\boldsymbol{x}) \boldsymbol{\mathcal{S}} \boldsymbol{\mathcal{B}}(\boldsymbol{x}) \right)^{-1}$ in Eq. (3.99) exists. Note that the matrix $\boldsymbol{\mathcal{S}}$ was assumed to be symmetric, but $\boldsymbol{\mathcal{S}}$ does not need to be positive definite to satisfy Eq. (3.101).

The matrix $\boldsymbol{\mathcal{B}}^g_{\boldsymbol{\mathcal{S}}}(\boldsymbol{x})$ is used to define the two complementary $n \times n$ projectors

$$\boldsymbol{\mathcal{P}}_{\boldsymbol{\mathcal{S}}}(\boldsymbol{x}) = \boldsymbol{\mathcal{B}}(\boldsymbol{x}) \boldsymbol{\mathcal{B}}^g_{\boldsymbol{\mathcal{S}}}(\boldsymbol{x}) = \boldsymbol{\mathcal{B}}(\boldsymbol{x}) \left(\boldsymbol{\mathcal{B}}^T(\boldsymbol{x}) \boldsymbol{\mathcal{S}} \boldsymbol{\mathcal{B}}(\boldsymbol{x}) \right)^{-1} \boldsymbol{\mathcal{B}}^T(\boldsymbol{x}) \boldsymbol{\mathcal{S}}, \tag{3.102}$$

$$\boldsymbol{\mathcal{Q}}_{\boldsymbol{\mathcal{S}}}(\boldsymbol{x}) = \boldsymbol{1} - \boldsymbol{\mathcal{P}}_{\boldsymbol{\mathcal{S}}}(\boldsymbol{x}). \tag{3.103}$$

The matrix $\boldsymbol{\mathcal{P}}_{\boldsymbol{\mathcal{S}}}(\boldsymbol{x})$ is idempotent but not symmetric,

$$\boldsymbol{\mathcal{P}}_{\boldsymbol{\mathcal{S}}}(\boldsymbol{x}) \boldsymbol{\mathcal{P}}_{\boldsymbol{\mathcal{S}}}(\boldsymbol{x}) = \boldsymbol{\mathcal{P}}_{\boldsymbol{\mathcal{S}}}(\boldsymbol{x}), \qquad \boldsymbol{\mathcal{P}}_{\boldsymbol{\mathcal{S}}}(\boldsymbol{x}) \neq \boldsymbol{\mathcal{P}}^T_{\boldsymbol{\mathcal{S}}}(\boldsymbol{x}), \tag{3.104}$$

and analogously for $\boldsymbol{\mathcal{Q}}_{\boldsymbol{\mathcal{S}}}(\boldsymbol{x})$. Using the solution Eq. (3.98) for the control in the controlled state equation (3.66) together with $\boldsymbol{x}(t) = \boldsymbol{x}_d(t)$ yields

$$\begin{aligned} \dot{\boldsymbol{x}}_d(t) - \boldsymbol{R} \left(\boldsymbol{x}_d(t) \right) &= \boldsymbol{\mathcal{B}} \left(\boldsymbol{x}_d(t) \right) \boldsymbol{u}(t) \\ &= \boldsymbol{\mathcal{P}}_{\boldsymbol{\mathcal{S}}} \left(\boldsymbol{x}_d(t) \right) \left(\dot{\boldsymbol{x}}_d(t) - \boldsymbol{R} \left(\boldsymbol{x}_d(t) \right) \right), \end{aligned} \tag{3.105}$$

and finally

$$\boldsymbol{\mathcal{Q}}_{\boldsymbol{\mathcal{S}}} \left(\boldsymbol{x}_d(t) \right) \left(\dot{\boldsymbol{x}}_d(t) - \boldsymbol{R} \left(\boldsymbol{x}_d(t) \right) \right) = \boldsymbol{0}. \tag{3.106}$$

Equation (3.106) looks very much like the constraint equation (3.80) found in the last section, but with a different projector $\boldsymbol{\mathcal{Q}}_{\boldsymbol{\mathcal{S}}}(\boldsymbol{x})$ instead of $\boldsymbol{\mathcal{Q}}(\boldsymbol{x})$. Additionally, the control signal Eq. (3.98) appears unequal from Eq. (3.79) obtained in the last section. It seems that, for the same problem, two unconnected control solutions were found. The first control solution in terms of $\boldsymbol{\mathcal{B}}^{+}$ and $\boldsymbol{\mathcal{Q}}$ is given by

$$\boldsymbol{u}_1(t) = \boldsymbol{\mathcal{B}}^{+}(\boldsymbol{x}_d(t))\left(\dot{\boldsymbol{x}}_d(t) - \boldsymbol{R}(\boldsymbol{x}_d(t))\right), \tag{3.107}$$

$$\boldsymbol{0} = \boldsymbol{\mathcal{Q}}(\boldsymbol{x}_d(t))\left(\dot{\boldsymbol{x}}_d(t) - \boldsymbol{R}(\boldsymbol{x}_d(t))\right), \tag{3.108}$$

while the second control solution in terms of $\boldsymbol{\mathcal{B}}^{g}_{\boldsymbol{\mathcal{S}}}$ and $\boldsymbol{\mathcal{Q}}_{\boldsymbol{\mathcal{S}}}$ is

$$\boldsymbol{u}_2(t) = \boldsymbol{\mathcal{B}}^{g}_{\boldsymbol{\mathcal{S}}}(\boldsymbol{x}_d(t))\left(\dot{\boldsymbol{x}}_d(t) - \boldsymbol{R}(\boldsymbol{x}_d(t))\right), \tag{3.109}$$

$$\boldsymbol{0} = \boldsymbol{\mathcal{Q}}_{\boldsymbol{\mathcal{S}}}(\boldsymbol{x}_d(t))\left(\dot{\boldsymbol{x}}_d(t) - \boldsymbol{R}(\boldsymbol{x}_d(t))\right). \tag{3.110}$$

The expressions $\boldsymbol{u}_1(t)$ and $\boldsymbol{u}_2(t)$ are identical for a matrix of weighting coefficients $\boldsymbol{\mathcal{S}} = \boldsymbol{1}$ but seem to disagree for $\boldsymbol{\mathcal{S}} \neq \boldsymbol{1}$.

However, the difference in $\boldsymbol{u}_2(t)$ and $\boldsymbol{u}_1(t)$ is deceptive. In fact, the expressions are identical, as is demonstrated in the following. Computing the difference between $\boldsymbol{u}_2(t)$ and $\boldsymbol{u}_1(t)$, multiplying by $\boldsymbol{\mathcal{B}}(\boldsymbol{x}_d(t))$, adding $\boldsymbol{0} = \boldsymbol{1} - \boldsymbol{1}$, and exploiting the constraint equations yields, see also Appendix A.2,

$$\begin{aligned}
\boldsymbol{\mathcal{B}}(\boldsymbol{x}_d(t))\left(\boldsymbol{u}_1(t) - \boldsymbol{u}_2(t)\right) &= \left(\boldsymbol{\mathcal{P}}(\boldsymbol{x}_d(t)) - \boldsymbol{\mathcal{P}}_{\boldsymbol{\mathcal{S}}}(\boldsymbol{x}_d(t))\right)\left(\dot{\boldsymbol{x}}_d(t) - \boldsymbol{R}(\boldsymbol{x}_d(t))\right) \\
&= \left(\boldsymbol{\mathcal{Q}}(\boldsymbol{x}_d(t)) - \boldsymbol{\mathcal{Q}}_{\boldsymbol{\mathcal{S}}}(\boldsymbol{x}_d(t))\right)\left(\dot{\boldsymbol{x}}_d(t) - \boldsymbol{R}(\boldsymbol{x}_d(t))\right) \\
&= \boldsymbol{0}.
\end{aligned} \tag{3.111}$$

Equation (3.111) implies that either $\boldsymbol{u}_1(t) = \boldsymbol{u}_2(t)$, or $\boldsymbol{u}_1(t) - \boldsymbol{u}_2(t)$ lies in the null space of $\boldsymbol{\mathcal{B}}(\boldsymbol{x}_d(t))$. Because $\boldsymbol{\mathcal{B}}(\boldsymbol{x})$ has full column rank for all $\boldsymbol{x}$ by assumption, the null space of $\boldsymbol{\mathcal{B}}(\boldsymbol{x})$ contains only the zero vector. In conclusion,

$$\boldsymbol{u}_1(t) = \boldsymbol{u}_2(t), \tag{3.112}$$

and, consequently, the control signal is unique and does not depend on the matrix of weighting coefficients $\boldsymbol{\mathcal{S}}$. Because identical control signals enforce identical controlled state trajectories $\boldsymbol{x}(t) = \boldsymbol{x}_d(t)$, the desired trajectories constrained by the different constraint equations (3.108) and (3.108) are identical as well.

In the framework of exactly realizable desired trajectories, the appearance of alternative projectors $\boldsymbol{\mathcal{P}}_{\boldsymbol{\mathcal{S}}}(\boldsymbol{x})$ and $\boldsymbol{\mathcal{Q}}_{\boldsymbol{\mathcal{S}}}(\boldsymbol{x})$ plays no role. Neither the control signal nor the controlled state trajectory depends on the matrix of weighting coefficients $\boldsymbol{\mathcal{S}}$. However, $\boldsymbol{\mathcal{P}}_{\boldsymbol{\mathcal{S}}}(\boldsymbol{x})$ and $\boldsymbol{\mathcal{Q}}_{\boldsymbol{\mathcal{S}}}(\boldsymbol{x})$ become important for the perturbative approach to trajectory tracking of arbitrary desired trajectories in Chapter 4. Note that the matrix $\boldsymbol{\mathcal{Q}}_{\boldsymbol{\mathcal{S}}}(\boldsymbol{x})$ is similar to the matrix $\boldsymbol{\mathcal{Q}}(\boldsymbol{x})$, i.e., there exists an invertible $n \times n$ matrix $\boldsymbol{\mathcal{T}}(\boldsymbol{x})$ such that

$$\boldsymbol{\mathcal{Q}}_{\boldsymbol{\mathcal{S}}}(\boldsymbol{x}) = \boldsymbol{\mathcal{T}}^{-1}(\boldsymbol{x})\,\boldsymbol{\mathcal{Q}}(\boldsymbol{x})\,\boldsymbol{\mathcal{T}}(\boldsymbol{x}). \tag{3.113}$$

This follows from the fact that both $\boldsymbol{\mathcal{Q}}_{\boldsymbol{\mathcal{S}}}(\boldsymbol{x})$ and $\boldsymbol{\mathcal{Q}}(\boldsymbol{x})$ are projectors of rank $n-p$. They have identical eigenvalues and, when diagonalized, identical diagonal forms $\boldsymbol{\mathcal{Q}}_D$. See also Appendix A.4 how to diagonalize the projectors $\boldsymbol{\mathcal{P}}(\boldsymbol{x})$ and $\boldsymbol{\mathcal{Q}}(\boldsymbol{x})$.

3.5. An exactly solvable example

A simple linear and exactly solvable example for optimal trajectory tracking is considered in this section. The rather clumsy exact solution is simplified by assuming a small regularization parameter $0 < \epsilon \ll 1$. A generalized perturbation expansion known as a singular perturbation expansion is necessary to obtain an approximation which is valid over the whole time domain $t_0 \leq t \leq t_1$. The purpose of analyzing this exact solution is three-fold. First, it serves as a pedagogical example displaying similar difficulties as the nonlinear system in Chapter 4. Second, it provides a consistency check for the analytical results of Chapter 4. Third, the impact of different terminal conditions on the exact solution is analyzed.

3.5.1. Problem and exact solution

Optimal trajectory tracking for a free particle with a finite regularization coefficient $\epsilon > 0$ is considered, see Example 3.1. For simplicity, a vanishing desired trajectory $\boldsymbol{x}_d(t) \equiv \mathbf{0}$ and zero initial time $t_0 = 0$ is assumed. The optimal control problem is to minimize the constrained functional

$$\mathcal{J}[\boldsymbol{x}(t), u(t)] = \frac{1}{2}\int_{t_0}^{t_1} \left((x(t))^2 + (y(t))^2\right) dt + \frac{\epsilon^2}{2}\int_{t_0}^{t_1} dt\,(u(t))^2 + \frac{\beta_1}{2}(x(t_1) - x_1)^2 + \frac{\beta_2}{2}(y(t_1) - y_1)^2, \tag{3.114}$$

subject to the system dynamics and initial conditions

$$\dot{x}(t) = y(t), \qquad \dot{y}(t) = u(t), \tag{3.115}$$

$$x(0) = x_0, \qquad y(0) = y_0. \tag{3.116}$$

Note that as long as $\boldsymbol{x}_0 \neq \mathbf{0}$, the desired trajectory $\boldsymbol{x}_d(t) \equiv \mathbf{0}$ does not comply with the initial conditions and is therefore not exactly realizable. Similarly, if $\boldsymbol{x}_1 \neq \mathbf{0}$, the desired trajectory $\boldsymbol{x}_d(t)$ does not comply with the terminal conditions for the state.

The co-state equation becomes

$$-\begin{pmatrix} \dot{\lambda}_x(t) \\ \dot{\lambda}_y(t) \end{pmatrix} = \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix}\begin{pmatrix} \lambda_x(t) \\ \lambda_y(t) \end{pmatrix} + \begin{pmatrix} x(t) \\ y(t) \end{pmatrix}, \tag{3.117}$$

with terminal condition

$$\lambda_x(t_1) = \beta_1(x(t_1) - x_1), \qquad \lambda_y(t_1) = \beta_2(y(t_1) - y_1). \tag{3.118}$$

The stationarity condition is

$$0 = \epsilon^2 u(t) + \lambda_y(t). \tag{3.119}$$

The state and co-state equations together with Eq. (3.119) can be cast in form of a 4×4 linear dynamical system

$$\begin{pmatrix} \dot{x}(t) \\ \dot{y}(t) \\ \dot{\lambda}_x(t) \\ \dot{\lambda}_y(t) \end{pmatrix} = \begin{pmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & -\epsilon^{-2} \\ -1 & 0 & 0 & 0 \\ 0 & -1 & -1 & 0 \end{pmatrix} \begin{pmatrix} x(t) \\ y(t) \\ \lambda_x(t) \\ \lambda_y(t) \end{pmatrix}. \tag{3.120}$$

Assuming the regularization parameter ϵ is restricted to the range $0 < \epsilon < \frac{1}{2}$, the four eigenvalues σ of the constant state matrix

$$\boldsymbol{\mathcal{A}} = \begin{pmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & -\epsilon^{-2} \\ -1 & 0 & 0 & 0 \\ 0 & -1 & -1 & 0 \end{pmatrix} \tag{3.121}$$

are real and given by

$$\sigma_{1,2,3,4} = \pm \frac{1}{\sqrt{2}\epsilon} \sqrt{1 \pm \sqrt{1 - 4\epsilon^2}}. \tag{3.122}$$

The exact solution for the state and co-state is given as a superposition of exponentials $\sim \exp(\sigma_i t)$, which is conveniently written as

$$\begin{aligned} \begin{pmatrix} x(t) \\ y(t) \\ \lambda_x(t) \\ \lambda_y(t) \end{pmatrix} = & \begin{pmatrix} \mathcal{H}_{11} \\ \mathcal{H}_{21} \\ \mathcal{H}_{31} \\ \mathcal{H}_{41} \end{pmatrix} \sinh(\kappa_1 t) + \begin{pmatrix} \mathcal{H}_{12} \\ \mathcal{H}_{22} \\ \mathcal{H}_{32} \\ \mathcal{H}_{42} \end{pmatrix} \cosh(\kappa_2 t) \\ & + \begin{pmatrix} \mathcal{H}_{13} \\ \mathcal{H}_{23} \\ \mathcal{H}_{33} \\ \mathcal{H}_{43} \end{pmatrix} \sinh(\kappa_1 t) + \begin{pmatrix} \mathcal{H}_{14} \\ \mathcal{H}_{24} \\ \mathcal{H}_{34} \\ \mathcal{H}_{44} \end{pmatrix} \sinh(\kappa_2 t). \end{aligned} \tag{3.123}$$

We introduced the abbreviations

$$\kappa_1 = \frac{\sqrt{1 - \sqrt{1 - 4\epsilon^2}}}{\sqrt{2}\epsilon}, \qquad \kappa_2 = \frac{\sqrt{\sqrt{1 - 4\epsilon^2} + 1}}{\sqrt{2}\epsilon}, \qquad \kappa_3 = \epsilon^2 \sqrt{1 - 4\epsilon^2}, \tag{3.124}$$

and the 4×4 matrix of coefficients $(\boldsymbol{\mathcal{H}})_{ij} = \mathcal{H}_{ij}$ given by

$$\boldsymbol{\mathcal{H}} = \begin{pmatrix} \frac{2C_2\epsilon^2 + x_0\left(\kappa_3 + \epsilon^2\right)}{2\kappa_3} & \frac{1}{2}\left(x_0 - \frac{\epsilon^2(2C_2 + x_0)}{\kappa_3}\right) & \frac{\kappa_3 y_0 - \epsilon^2(2C_1 + y_0)}{2\kappa_1\kappa_3} & \frac{2C_1\epsilon^2 + y_0\left(\kappa_3 + \epsilon^2\right)}{2\kappa_2\kappa_3} \\ \frac{1}{2}\left(y_0 - \frac{\epsilon^2(2C_1 + y_0)}{\kappa_3}\right) & \frac{2C_1\epsilon^2 + y_0\left(\kappa_3 + \epsilon^2\right)}{2\kappa_3} & \frac{C_2\left(\epsilon^2 - \kappa_3\right) + 2x_0\epsilon^4}{2\kappa_1\kappa_3\epsilon^2} & -\frac{C_2\left(\kappa_3 + \epsilon^2\right) + 2x_0\epsilon^4}{2\kappa_2\kappa_3\epsilon^2} \\ \frac{C_1\left(\kappa_3 + \epsilon^2\right) + 2y_0\epsilon^4}{2\kappa_3} & \frac{C_1\left(\kappa_3 - \epsilon^2\right) - 2y_0\epsilon^4}{2\kappa_3} & -\frac{2C_2\epsilon^2 + x_0\left(\kappa_3 + \epsilon^2\right)}{2\kappa_1\kappa_3} & \frac{2C_2\epsilon^2 + x_0\left(\epsilon^2 - \kappa_3\right)}{2\kappa_2\kappa_3} \\ \frac{C_2\left(\kappa_3 - \epsilon^2\right) - 2x_0\epsilon^4}{2\kappa_3} & \frac{C_2\left(\kappa_3 + \epsilon^2\right) + 2x_0\epsilon^4}{2\kappa_3} & \frac{-\kappa_3(C_1 + y_0) + \epsilon^2(C_1 + y_0) - 2y_0\epsilon^4}{2\kappa_1\kappa_3} & -\frac{2C_1\kappa_2^2\epsilon^4 + y_0\left(\kappa_3 - 2\epsilon^4 + \epsilon^2\right)}{2\kappa_2\kappa_3} \end{pmatrix}. \tag{3.125}$$

The constants C_1 and C_2 are very long expressions and not important for the subsequent discussion.

Although the solution Eq. (3.123) is unhandy, it can be studied relatively easily with the computer algebra system Mathematica (Wolfram Research, Inc., 2014). The next section compares Eq. (3.123) with exact solutions to slightly different optimization problems. These problems differ from Eq. (3.114) only in the terminal conditions.

3.5.2. Different terminal conditions

The last section discussed the case of penalized terminal conditions leading to terminal conditions for the co-state,

$$\lambda_x(t_1) = \beta_1 (x(t_1) - x_1), \qquad \lambda_y(t_1) = \beta_2 (y(t_1) - y_1). \tag{3.126}$$

The weighting coefficients $\beta_{1/2} \geq 0$ quantify the cost for deviating from the desired terminal state $\boldsymbol{x}_1$. With increasing value of $\beta_{1/2}$, the cost is increasing. In the limit $\beta_{1/2} \to \infty$, the terminal co-state can be finite only for sharp terminal conditions

$$x(t_1) = x_1, \qquad y(t_1) = y_1. \tag{3.127}$$

This limit can only exist if the system is controllable. Controllability for mechanical control systems in one spatial dimension, including the free particle discussed here, was proven in Section 2.4.

These considerations lead to the following conjecture. Minimizing the functional

$$\mathcal{J}[\boldsymbol{x}(t), u(t)] = \frac{1}{2}\int_{t_0}^{t_1} \left((x(t))^2 + (y(t))^2\right) dt + \frac{\epsilon^2}{2}\int_{t_0}^{t_1} dt\,(u(t))^2 + \frac{\beta_1}{2}(x(t_1) - x_1)^2 + \frac{\beta_2}{2}(y(t_1) - y_1)^2 \tag{3.128}$$

subject to

$$\dot{x}(t) = y(t), \qquad \dot{y}(t) = u(t), \qquad x(0) = x_0, \qquad y(0) = y_0, \tag{3.129}$$

and subsequently applying the limit $\beta_{1/2} \to \infty$, is equivalent to the minimization of

$$\mathcal{J}[\boldsymbol{x}(t), u(t)] = \frac{1}{2}\int_{t_0}^{t_1} \left((x(t))^2 + (y(t))^2\right) dt + \frac{\epsilon^2}{2}\int_{t_0}^{t_1} dt\,(u(t))^2 \tag{3.130}$$

subject to

$$\dot{x}(t) = y(t), \qquad \dot{y}(t) = u(t), \tag{3.131}$$

$$x(0) = x_0, \qquad y(0) = y_0, \qquad x(t_1) = x_1, \qquad y(t_1) = y_1. \tag{3.132}$$

This conjecture is confirmed by computing the limit $\beta_{1/2} \to \infty$ of the exact solution, Eq. (3.123), and comparing it with the exact solution to the corresponding problem with sharp terminal conditions Eq. (3.132).

Similarly, we confirm that the limit $\beta_{1/2} \to 0$ of Eq. (3.123) is equivalent to the optimization problem with free terminal conditions, i.e., to minimizing the functional

$$\mathcal{J}\left[\boldsymbol{x}\left(t\right), u\left(t\right)\right] = \frac{1}{2}\int_{t_0}^{t_1}\left(\left(x\left(t\right)\right)^2 + \left(y\left(t\right)\right)^2\right)dt + \frac{\epsilon^2}{2}\int_{t_0}^{t_1} dt\left(u\left(t\right)\right)^2 \tag{3.133}$$

subject to

$$\dot{x}\left(t\right) = y\left(t\right), \qquad \dot{y}\left(t\right) = u\left(t\right), \qquad x\left(0\right) = x_0, \qquad y\left(0\right) = y_0. \tag{3.134}$$

In conclusion, the task of minimizing Eq. (3.114), with exact solution Eq. (3.123), is the most general way to formulate the terminal condition. All other cases can be generated by applying the appropriate limits of the weighting coefficients $\beta_{1/2}$. This insight enables an efficient perturbative treatment of the general nonlinear problem in Chapter 4. However, the perturbative approach uses the regularization parameter ϵ as the small parameter, which corresponds to investigating the limit $\epsilon \to 0$. As demonstrated in the next section, the limit $\epsilon \to 0$ is not without difficulties. The subtle question arises if applying the limit $\epsilon \to 0$ commutes with the limit $\beta_{1/2} \to 0$.

3.5.3. Approximating the exact solution

A small regularization parameter $0 < \epsilon \ll 1$ is assumed to approximate the exact solution. A regular expansion in form of a power series in ϵ results in an approximation which is not uniformly valid over the entire time interval. Phenomenologically, this non-uniformity manifests in the appearance of *boundary layers* at the beginning and end of the time interval for small ϵ. Uniformly valid approximation are obtained for a series expansion in form of an asymptotic series. This procedure is known as a singular perturbation expansion. For a detailed account of asymptotic series, the difference between singular and regular perturbation theory, and boundary layers see the excellent book (Bender and Orszag, 2010) and also (Johnson, 2004).

3.5.3.1. Inner and outer limits

The solution Eq. (3.123) is rather intimidating and clumsy and not very useful for subsequent computations. Here, the leading order approximation to Eq. (3.123) for small ϵ is obtained. The perturbation expansion also involves expansion of the eigenvalues σ_i. Due to the appearance of $1/\epsilon$ in the eigenvalues σ_i, care has to be

taken when considering the limit $\epsilon \to 0$. Note that $\kappa_{1,2}$ behave as

$$\kappa_1 = \frac{\sqrt{1-\sqrt{1-4\epsilon^2}}}{\sqrt{2}\epsilon} = 1 + \frac{\epsilon^2}{2} + \mathcal{O}\left(\epsilon^3\right), \tag{3.135}$$

$$\kappa_2 = \frac{\sqrt{\sqrt{1-4\epsilon^2}+1}}{\sqrt{2}\epsilon} = \frac{1}{\epsilon} - \frac{\epsilon}{2} + \mathcal{O}\left(\epsilon^3\right), \tag{3.136}$$

such that the exponential terms contained in Eq. (3.123) are of the form

$$\exp\left(-\kappa_2 t\right) \approx \exp\left(-\frac{t}{\epsilon}\right). \tag{3.137}$$

$$\exp\left(-\kappa_1 t\right) \approx \exp\left(-t\right), \tag{3.138}$$

While the limit for Eq. (3.138) can safely be applied independent of the value of t, the limit for Eq. (3.137) depends on the actual value of t. The result is

$$\lim_{\epsilon\to 0} \exp\left(-\frac{1}{\epsilon}t\right) = \begin{cases} 0, & t > 0, \\ 1, & t = 0, \end{cases} \tag{3.139}$$

i.e., the exact solution (3.123) for $\epsilon = 0$ is discontinuous and jumps at the left end of the time domain. For small but finite ϵ, such a behavior results in the appearance boundary layers. Analytically, these can be resolved by rescaling time t appropriately with the small parameter ϵ. Because the solution contains (for small ϵ) also exponential terms of the form $\exp\left(-\left(t-t_1\right)/\epsilon\right)$, a similar boundary layer is expected at the right end point $t = t_1$ of the time domain.

First, consider the limit $\epsilon \to 0$ in an interior point $0 < t < t_1$ of the time domain. This limit is called the *outer limit* and is denoted with index O,

$$\begin{pmatrix} x_O\left(t\right) \\ y_O\left(t\right) \\ \lambda_{x,O}\left(t\right) \\ \lambda_{y,O}\left(t\right) \end{pmatrix} = \lim_{\epsilon\to 0} \begin{pmatrix} x\left(t\right) \\ y\left(t\right) \\ \lambda_x\left(t\right) \\ \lambda_y\left(t\right) \end{pmatrix}. \tag{3.140}$$

To obtain this limit by hand is very tedious due to the complexity of the exact solution. We rely on the capabilities of the computer algebra system Mathematica and simply state the result,

$$\begin{pmatrix} x_O\left(t\right) \\ y_O\left(t\right) \\ \lambda_{x,O}\left(t\right) \\ \lambda_{y,O}\left(t\right) \end{pmatrix} = \begin{pmatrix} \frac{\beta_1 x_1}{\kappa}\sinh\left(t\right) + \frac{x_0}{\kappa}\left(\cosh\left(t-t_1\right) - \beta_1\sinh\left(t-t_1\right)\right) \\ \frac{\beta_1 x_1}{\kappa}\cosh\left(t\right) + \frac{x_0}{\kappa}\left(\sinh\left(t-t_1\right) - \beta_1\cosh\left(t-t_1\right)\right) \\ \frac{1}{\kappa}\cosh\left(t\right)\left(x_0\left(\beta_1\cosh\left(t_1\right) + \sinh\left(t_1\right)\right) - \beta_1 x_1\right) - x_0\sinh\left(t\right) \\ 0 \end{pmatrix}, \tag{3.141}$$

with the abbreviation

$$\kappa = \cosh(t_1) + \beta_1 \sinh(t_1). \tag{3.142}$$

Note that this solution does not depend on the initial and terminal points y_0 and y_1! Furthermore, the outer limit $y_O(t)$ does not obey the initial condition $y(0) = y_0$ because

$$y_O(0) = \frac{\beta_1 x_1}{\kappa} - \frac{x_0}{\kappa}(\sinh(t_1) + \beta_1 \cosh(t_1)) \neq y_0. \tag{3.143}$$

To obtain the behavior of the exact solution (3.123) near to $t = t_0 = 0$, the time t is rescaled and a new time scale is introduced as

$$\tau_L = (t - t_0)/\epsilon = t/\epsilon. \tag{3.144}$$

The corresponding limits of Eq. (3.123), valid for times t on the order of ϵ, $t \sim \epsilon$, are called *left inner limits* and denoted by

$$\begin{pmatrix} X(\tau_L) \\ Y_L(\tau_L) \\ \Lambda_{x,L}(\tau_L) \\ \Lambda_{y,L}(\tau_L) \end{pmatrix} = \lim_{\epsilon \to 0} \begin{pmatrix} x(\epsilon\tau_L) \\ y(\epsilon\tau_L) \\ \lambda_x(\epsilon\tau_L) \\ \lambda_y(\epsilon\tau_L) \end{pmatrix}. \tag{3.145}$$

With the help of Mathematica, we obtain

$$\begin{pmatrix} X_L(\tau_L) \\ Y_L(\tau_L) \\ \Lambda_{x,L}(\tau_L) \\ \Lambda_{y,L}(\tau_L) \end{pmatrix} = \begin{pmatrix} x_0 \\ \frac{(1 - e^{-\tau_L})}{\kappa}(\beta_1 x_1 - x_0(\beta_1 \cosh(t_1) + \sinh(t_1))) + y_0 e^{-\tau_L} \\ \frac{\beta_1 \operatorname{csch}(t_1)(x_0 \cosh(t_1) - x_1) + x_0}{\beta_1 + \coth(t_1)} \\ 0 \end{pmatrix}. \tag{3.146}$$

Note that the solution $Y_L(\tau_L)$ satisfies the appropriate initial condition

$$Y_L(0) = y(0) = y_0. \tag{3.147}$$

A similar procedure is applied for the *right inner limit* by introducing an appropriate time scale as

$$\tau_R = (t_1 - t)/\epsilon. \tag{3.148}$$

Solutions valid for times t close to the end of the time interval, $(t_1 - t) \sim \epsilon$, are denoted by

$$\begin{pmatrix} X_R(\tau_R) \\ Y_R(\tau_R) \\ \Lambda_{x,R}(\tau_R) \\ \Lambda_{y,R}(\tau_R) \end{pmatrix} = \lim_{\epsilon \to 0} \begin{pmatrix} x(t_1 - \epsilon\tau_R) \\ y(t_1 - \epsilon\tau_R) \\ \lambda_x(t_1 - \epsilon\tau_R) \\ \lambda_y(t_1 - \epsilon\tau_R) \end{pmatrix}. \tag{3.149}$$

With the help of Mathematica, we obtain

$$\begin{pmatrix} X_R(\tau_R) \\ Y_R(\tau_R) \\ \Lambda_{x,R}(\tau_R) \\ \Lambda_{y,R}(\tau_R) \end{pmatrix} = \begin{pmatrix} \frac{1}{\kappa}(x_0 + x_1\beta_1 \sinh(t_1)) \\ y_1 e^{-\tau_R} - \frac{\beta_1}{\kappa}(1 - e^{-\tau_R})(x_0 - x_1 \cosh(t_1)) \\ \frac{\beta_1}{\kappa}(x_0 - x_1 \cosh(t_1)) \\ 0 \end{pmatrix}. \tag{3.150}$$

Fig. 3.1 shows the exact solution for the state variable $y(t)$ (red line) together with the corresponding outer and both inner limits (dashed lines) for a relatively large value of $\epsilon = 1/10$. While the outer limit provides an approximation inside the domain but fails at the beginning and end of the time domain, the left and right inner limits approximate these regions quite well. The idea is now to combine all limits in a single composite solution and obtain an approximation which is uniformly valid over the entire time interval.

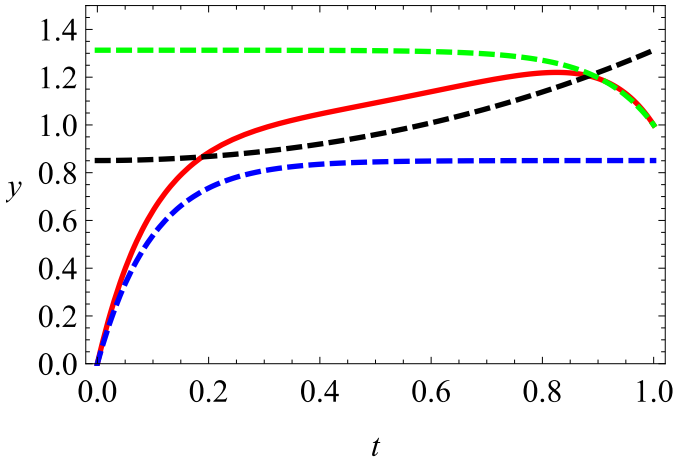

Figure 3.1.: Exact solution for the state component $y(t)$ (red solid line) for $\epsilon = 0.1$ shows the appearance of boundary layers at the beginning and end of the time domain. The black dashed line shows the outer approximation $y_O(t)$ valid in the bulk of the time domain. The blue and green dashed lines shows that the left and right inner approximations given by $Y_L(t/\epsilon)$ and $Y_R((t_1 - t)/\epsilon)$ are valid close to the initial and terminal time, respectively, and resolve the boundary layers. A combination of all three solutions is necessary to yield an approximation which is uniformly valid over the whole time interval and also satisfies the initial and terminal conditions.

3.5.3.2. Matching and composite solution

For a composition of the inner and outer limits to a single and uniformly valid approximation, certain *matching conditions* must be satisfied. If the matching conditions are violated, other scalings of time t with the small parameter ϵ exist and

cannot be neglected in the composite solution. In general, several scaling regimes for the inner solutions are possible. A larger variety of scalings may lead to more complicated structures such as *nested boundary layers*, *interior boundary layers* or *super sharp boundary layers* (Bender and Orszag, 2010).

The matching conditions at the left end of the time domain are

$$O_L = \lim_{t\to 0} \begin{pmatrix} x_O(t) \\ y_O(t) \\ \lambda_{x,O}(t) \\ \lambda_{y,O}(t) \end{pmatrix} = \lim_{\tau_L\to\infty} \begin{pmatrix} X(\tau_L) \\ Y_L(\tau_L) \\ \Lambda_{x,L}(\tau_L) \\ \Lambda_{y,L}(\tau_L) \end{pmatrix}. \tag{3.151}$$

Computing both limits yields identical results, and the matching conditions are satisfied. The *left overlap* O_L is obtained as

$$O_L = \begin{pmatrix} x_0 \\ \frac{1}{\kappa}\left(\beta_1 x_1 - x_0\left(\beta_1\cosh(t_1) + \sinh(t_1)\right)\right) \\ \frac{\beta_1 \operatorname{csch}(t_1)\left(x_0\cosh(t_1) - x_1\right) + x_0}{\beta_1 + \coth(t_1)} \\ 0 \end{pmatrix}. \tag{3.152}$$

The matching conditions at the right end of the time domain are

$$O_R = \lim_{t\to t_1} \begin{pmatrix} x_O(t) \\ y_O(t) \\ \lambda_{x,O}(t) \\ \lambda_{y,O}(t) \end{pmatrix} = \lim_{\tau_R\to\infty} \begin{pmatrix} X_R(\tau_R) \\ Y_R(\tau_R) \\ \Lambda_{x,R}(\tau_R) \\ \Lambda_{y,R}(\tau_R) \end{pmatrix}. \tag{3.153}$$

The *right overlap* O_R is obtained as

$$O_R = \begin{pmatrix} \frac{1}{\kappa}\left(x_0 + x_1\beta_1\sinh(t_1)\right) \\ -\frac{\beta_1}{\kappa}\left(x_0 - x_1\cosh(t_1)\right) \\ \frac{\beta_1}{\kappa}\left(x_0 - x_1\cosh(t_1)\right) \\ 0 \end{pmatrix}. \tag{3.154}$$

In conclusion, both matching conditions are satisfied.

Knowing all inner and outer approximations, they are combined to a *composite solution* uniformly valid on the entire time domain. This is done by adding up all

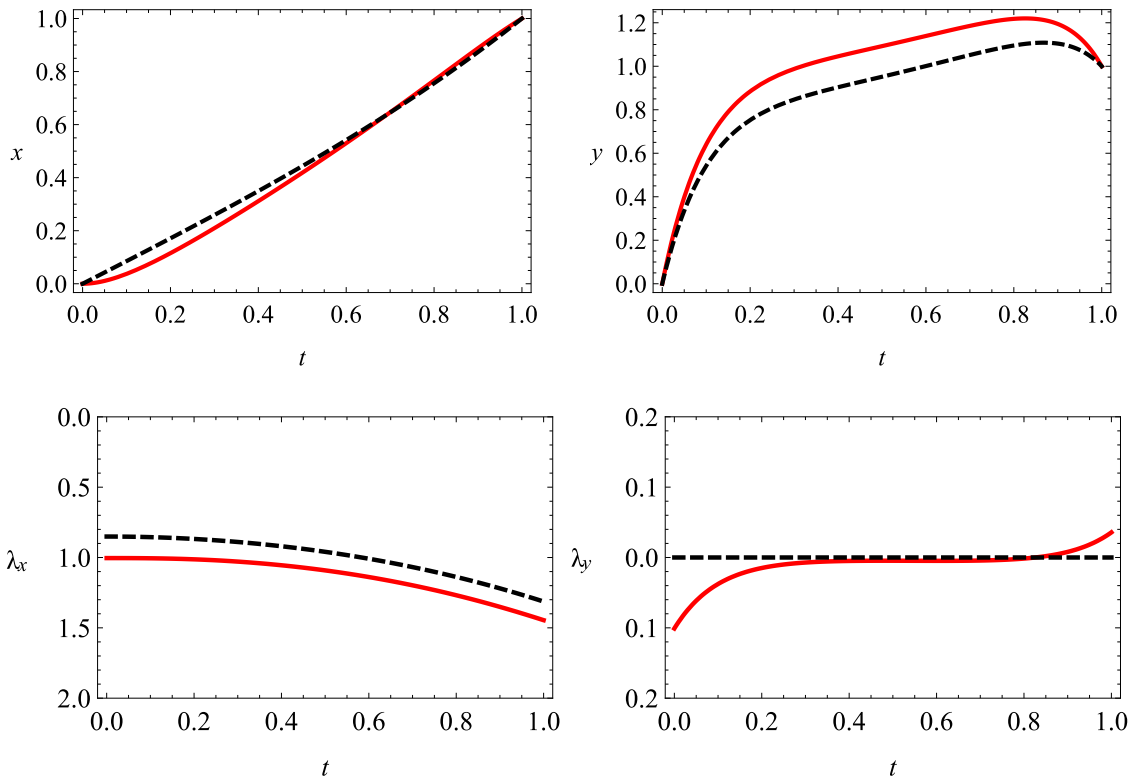

Figure 3.2.: The exact solutions (red solid line) for the states $x(t)$ (top left), $y(t)$ (top right) and co-states $\lambda_x(t)$ (bottom left) and $\lambda_y(t)$ (bottom right) for $\epsilon = 0.1$. The black dashed lines show the approximate composite solutions obtained by a singular perturbation expansion.

inner and outer solutions and subtracting the overlaps (Bender and Orszag, 2010),

$$\begin{pmatrix} x_{\text{comp}}(t) \\ y_{\text{comp}}(t) \\ \lambda_{x,\text{comp}}(t) \\ \lambda_{y,\text{comp}}(t) \end{pmatrix} = \begin{pmatrix} x_O(t) \\ y_O(t) \\ \lambda_{x,O}(t) \\ \lambda_{y,O}(t) \end{pmatrix} + \begin{pmatrix} X_L(t/\epsilon) \\ Y_L(t/\epsilon) \\ \Lambda_{x,L}(t/\epsilon) \\ \Lambda_{y,L}(t/\epsilon) \end{pmatrix} + \begin{pmatrix} X_R((t_1-t)/\epsilon) \\ Y_R((t_1-t)/\epsilon) \\ \Lambda_{x,R}((t_1-t)/\epsilon) \\ \Lambda_{y,R}((t_1-t)/\epsilon) \end{pmatrix} - O_L - O_R. \tag{3.155}$$

Finally, the composite solution is given by

$$\begin{aligned} y_{\text{comp}}(t) = {} & \frac{1}{\kappa}\left(\beta_1 x_1 \cosh(t) + x_0\left(\sinh(t-t_1) - \beta_1\cosh(t-t_1)\right)\right) \\ & + \frac{1}{\kappa}e^{-(t_1-t)/\epsilon}\left(\beta_1\left(y_1\sinh(t_1)+x_0\right) + \cosh(t_1)\left(y_1-\beta_1 x_1\right)\right) \\ & + \frac{1}{\kappa}e^{-t/\epsilon}\left(\sinh(t_1)\left(x_0+\beta_1 y_0\right) + \cosh(t_1)\left(\beta_1 x_0 + y_0\right) - \beta_1 x_1\right), \end{aligned} \tag{3.156}$$

and

$$\begin{pmatrix} x_{\text{comp}}(t) \\ \lambda_{x,\text{comp}}(t) \\ \lambda_{y,\text{comp}}(t) \end{pmatrix} = \begin{pmatrix} \frac{\beta_1 x_1}{\kappa}\sinh(t) + \frac{x_0}{\kappa}\left(\cosh(t-t_1) - \beta_1 \sinh(t-t_1)\right) \\ \frac{\cosh(t)}{\kappa}\left(x_0\left(\beta_1\cosh(t_1) + \sinh(t_1)\right) - \beta_1 x_1\right) - x_0\sinh(t) \\ 0 \end{pmatrix}. \tag{3.157}$$

This is the leading order approximation as $\epsilon \to 0$ of the exact solution Eq. (3.123). Note that this approximate solution depends on the small parameter ϵ itself, as it is generally the case for singular perturbation expansions. Thus, the leading order composite solution is not simply given by the limit

$$\begin{pmatrix} x_{\text{comp}}(t) \\ y_{\text{comp}}(t) \\ \lambda_{x,\text{comp}}(t) \\ \lambda_{y,\text{comp}}(t) \end{pmatrix} \neq \lim_{\epsilon\to 0} \begin{pmatrix} x(t) \\ y(t) \\ \lambda_x(t) \\ \lambda_y(t) \end{pmatrix}, \tag{3.158}$$

with $\left(x(t),\ \ y(t),\ \ \lambda_x(t),\ \ \lambda_y(t) \right)^T$ denoting the exact solution. To distinguish the operation "obtain the leading order contribution" from the limit $\epsilon \to 0$, the notation

$$\begin{pmatrix} x_{\text{comp}}(t) \\ y_{\text{comp}}(t) \\ \lambda_{x,\text{comp}}(t) \\ \lambda_{y,\text{comp}}(t) \end{pmatrix} = \begin{pmatrix} x(t) \\ y(t) \\ \lambda_x(t) \\ \lambda_y(t) \end{pmatrix} + \text{h.o.t.} \tag{3.159}$$

is adopted. Here, h.o.t. stands for "higher order terms", and means that all higher order contributions are neglected as $\epsilon \to 0$.

The composite solutions for all state and co-state variables (black dashed lines) are compared with the corresponding exact analytical solutions (red solid lines) in two figures. For a relatively large value of $\epsilon = 1/10$ the solutions agree at least qualitatively, see Fig. 3.2. For the smaller values of $\epsilon = 1/40$ the exact solution is approximated quite well. The boundary layers of width ϵ displayed by state component y become steeper with decreasing ϵ and degenerate to discontinuous jumps for $\epsilon = 0$.

3.5.3.3. Different end point conditions

Interestingly, the composite solution (3.156) and (3.157) does not depend on the weighting coefficient β_2. This can be explained as follows. Note that the co-state λ_y has to satisfy the terminal condition

$$\lambda_y(t_1) = \beta_2\left(y(t_1) - y_1\right), \tag{3.160}$$

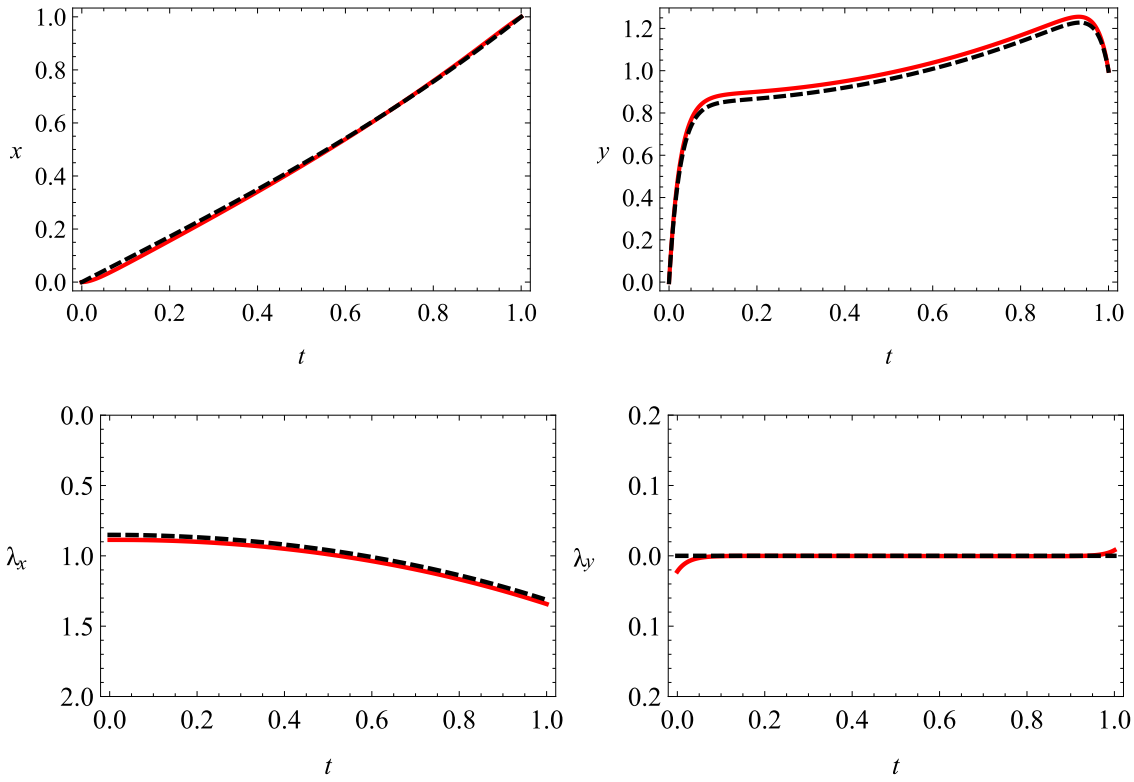

Figure 3.3.: Same as in Fig. 3.2 but for a smaller value of $\epsilon = 1/40 = 0.025$. The agreement between exact (red solid line) and approximate solution (black dashed line) becomes better for smaller values of ϵ while the slopes of the boundary layers exhibited by the state component $y(t)$ steepen.

as well as the stationarity condition

$$0 = \epsilon^2 u(t) + \lambda_y(t). \tag{3.161}$$

Clearly, a finite control signal $u(t) \neq 0$ together with $\epsilon = 0$ implies $\lambda_y(t) \equiv 0$. Thus, the terminal condition Eq. (3.160) can only be satisfied if

$$y(t_1) = y_1. \tag{3.162}$$

In the limit $\epsilon \to 0$, the state variable y satisfies the sharp terminal condition. Consequently, the parameter β_2 drops out of the equations, rendering state, co-state, and control signal independent of β_2.

A different question is how the composite solutions Eqs. (3.156) and (3.157) behave for $\beta_1 \to \infty$ and $\beta_1 = 0$. Due to the subtle nature of the leading order approximation as $\epsilon \to 0$, it is not immediately clear if computing the leading order commutes with the limits $\beta_1 \to \infty$ and $\beta_1 \to 0$. Comparison of the leading order approximations of the exact solutions for $\beta_1 \to \infty$ and $\beta_1 = 0$ with the limits $\beta_1 \to \infty$ and $\beta_1 \to 0$ of the

leading order composite solution (3.156) and (3.157) confirms that these operations commute. In other words, computing first

$$\begin{pmatrix} x^{\infty}(t) \\ y^{\infty}(t) \\ \lambda_x^{\infty}(t) \\ \lambda_y^{\infty}(t) \end{pmatrix} = \lim_{\beta_1 \to \infty} \begin{pmatrix} x(t) \\ y(t) \\ \lambda_x(t) \\ \lambda_y(t) \end{pmatrix} \tag{3.163}$$

followed by a leading order approximation as $\epsilon \to 0$ leads to the same result as computing the limit $\beta_1 \to \infty$ of the composite solution (3.156) and (3.157). This identity is expressed as

$$\lim_{\beta_1 \to \infty} \begin{pmatrix} x_{\text{comp}}(t) \\ \lambda_{x,\text{comp}}(t) \\ \lambda_{y,\text{comp}}(t) \end{pmatrix} = \begin{pmatrix} x^{\infty}(t) \\ y^{\infty}(t) \\ \lambda_x^{\infty}(t) \\ \lambda_y^{\infty}(t) \end{pmatrix} + \text{h.o.t.}. \tag{3.164}$$

An analogous result is valid for the limit $\beta_1 \to 0$. Let the exact result for $\beta_1 = 0$ be defined by

$$\begin{pmatrix} x^{0}(t) \\ y^{0}(t) \\ \lambda_x^{0}(t) \\ \lambda_y^{0}(t) \end{pmatrix} = \lim_{\beta_1 \to 0} \begin{pmatrix} x(t) \\ y(t) \\ \lambda_x(t) \\ \lambda_y(t) \end{pmatrix}. \tag{3.165}$$

Then the following identity holds

$$\lim_{\beta_1 \to 0} \begin{pmatrix} x_{\text{comp}}(t) \\ \lambda_{x,\text{comp}}(t) \\ \lambda_{y,\text{comp}}(t) \end{pmatrix} = \begin{pmatrix} x^{0}(t) \\ y^{0}(t) \\ \lambda_x^{0}(t) \\ \lambda_y^{0}(t) \end{pmatrix} + \text{h.o.t.}. \tag{3.166}$$

In summary, for the leading order approximation as $\epsilon \to 0$, it is sufficient to study the problem

$$\mathcal{J}[\boldsymbol{x}(t), u(t)] = \frac{1}{2} \int_{t_0}^{t_1} \left((x(t))^2 + (y(t))^2 \right) dt + \frac{\epsilon^2}{2} \int_{t_0}^{t_1} dt \, (u(t))^2 + \frac{\beta_1}{2} (x(t_1) - x_1)^2, \tag{3.167}$$

subject to

$$\dot{x}(t) = y(t), \quad \dot{y}(t) = u(t), \quad x(0) = x_0, \quad y(0) = y_0, \quad y(t_1) = y_1. \tag{3.168}$$

All other leading order approximations for sharp or free terminal conditions can be generated from the solution to Eqs. (3.167) and (3.168) by applying the limits $\beta_1 \to \infty$ and $\beta_1 = 0$.

3.5.3.4. Solution for the control signal

To obtain simpler expressions, the limit $\beta_1 \to \infty$ together with $x_0 = y_0 = 0$ is assumed in the remainder of this section.

It remains to find an approximation for the control signal $u(t)$ as follows. One way is to analyze the exact solution for the co-state $\lambda_y(t)$ together with Eq. (3.119) and perform the singular perturbation expansion to find the leading order approximation for $u(t)$. Instead, here the control signal is derived from the composite solution for state and co-state. Note that Eq. (3.119) is satisfied to the lowest order in ϵ also for the composite solution because $\lambda_{y,\mathrm{comp}}(t) = 0$ vanishes to leading order in ϵ. However, to compute $u(t)$ from Eq. (3.119) requires the knowledge of higher order contributions to $\lambda_{y,\mathrm{comp}}(t)$. As an alternative, Eq. (3.115) is utilized to obtain

$$\begin{aligned} u_{\mathrm{comp}}(t) &= \dot{y}_{\mathrm{comp}}(t) \\ &= x_1 \mathrm{csch}(t_1) \sinh(t) + x_1 \mathrm{csch}(t_1) \frac{e^{-\frac{t}{\epsilon}}}{\epsilon} + \frac{e^{-\frac{t_1 - t}{\epsilon}}}{\epsilon} \left(y_1 - x_1 \coth(t_1)\right). \end{aligned} \tag{3.169}$$

Similar to the composite solution for the state $y(t)$, the approximate control signal also contains two boundary layers, one at $t = 0$ and a second at $t = t_1$. The outer limit of Eq. (3.169), valid for times $t_0 < t < t_1$ is

$$u_O(t) = \lim_{\epsilon \to 0} u_{\mathrm{comp}}(t) = x_1 \mathrm{csch}(t_1) \sinh(t). \tag{3.170}$$

For the left and right inner limits, rescaled time scales $\tau_L = t/\epsilon$ and $\tau_R = (t_1 - t)/\epsilon$ are introduced. The corresponding leading order solutions are denoted by $U_L(\tau_L)$ and $U_R(\tau_R)$, respectively. Because of the factor $1/\epsilon$ in Eq. (3.169), it is impossible to apply the limit $\epsilon \to 0$. Instead, the leading order contribution is computed as

$$\begin{aligned} u(t) &= u(t_0 + \epsilon\tau_L) = u_{\mathrm{comp}}(t_0 + \epsilon\tau_L) + \text{h.o.t.} \\ &= U_L(\tau_L) + \text{h.o.t.}, \end{aligned} \tag{3.171}$$

$$\begin{aligned} u(t) &= u(t_1 - \epsilon\tau_R) = u_{\mathrm{comp}}(t_1 - \epsilon\tau_R) + \text{h.o.t.} \\ &= U_R(\tau_R) + \text{h.o.t.}. \end{aligned} \tag{3.172}$$

All contributions of higher order in ϵ are neglected. This procedure yields

$$U_L(\tau_L) = x_1 \mathrm{csch}(t_1) \frac{e^{-\tau_L}}{\epsilon}, \tag{3.173}$$

$$U_R(\tau_R) = x_1 + \frac{e^{-\tau_R}}{\epsilon} \left(y_1 - x_1 \coth(t_1)\right). \tag{3.174}$$

Finally, the matching conditions

$$u_O(0) = 0 = \lim_{\tau_L \to \infty} U_L(\tau_L), \tag{3.175}$$

$$u_O(t_1) = x_1 = \lim_{\tau_R \to \infty} U_R(\tau_R), \tag{3.176}$$

are satisfied, confirming that three scaling regions are sufficient for the composite solution. Hence, Eq. (3.169) is the leading order approximation to the exact solution for the control signal. Equation (3.169) (black dashed line) is compared with the exact solution (red solid line) for two different values of ϵ in Fig. 3.4.

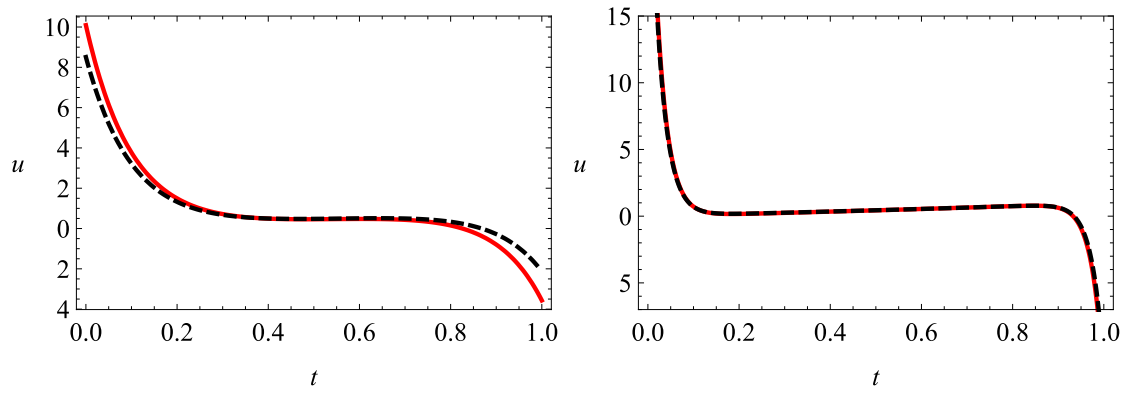

Figure 3.4.: Comparison of the exact control signal (red solid line) and its approximation (black dashed line) for two different values of the small parameter ϵ, $\epsilon = 1/10$ (left) and $\epsilon = 1/40$ (right). For ϵ approaching zero, the steeply rising boundary layers will approach infinity, leading to delta-like kicks located at the beginning and end of the time domain.

3.5.3.5. Exact solution for $\epsilon = 0$

Having obtained approximate solutions for small but finite $\epsilon > 0$, it becomes possible to understand what exactly happens for $\epsilon = 0$. The composite solution for the state component $y(t)$ for $\beta_1 \to \infty$ and initial conditions $x_0 = y_0 = 0$ is given by

$$\begin{aligned} y_{\text{comp}}(t) = {} & x_1 \text{csch}(t_1) \cosh(t) - x_1 \text{csch}(t_1)\, e^{-t/\epsilon} \\ & + e^{-(t_1-t)/\epsilon} \left(y_1 - x_1 \coth(t_1)\right). \end{aligned} \tag{3.177}$$

The exact solution for $\epsilon = 0$ is obtained by computing the limit

$$\lim_{\epsilon \to 0} y(t) = \lim_{\epsilon \to 0} y_{\text{comp}}(t) = \begin{cases} 0, & t = 0, \\ x_1 \text{csch}(t_1) \cosh(t), & 0 < t < t_1, \\ y_1, & t = t_1. \end{cases} \tag{3.178}$$

Using the Kronecker delta defined as

$$\delta_{a,b} = \begin{cases} 1, & a = b, \\ 0, & a \neq b, \end{cases} \tag{3.179}$$

Eq. (3.178) can be written in the form

$$\lim_{\epsilon\to 0} y(t) = x_1 \mathrm{csch}(t_1)\cosh(t) - x_1 \mathrm{csch}(t_1)\,\delta_{t,0} + (y_1 - x_1 \coth(t_1))\,\delta_{t,t_1}. \tag{3.180}$$

The control signal $u(t)$ for $\epsilon = 0$ is

$$\lim_{\epsilon\to 0} u(t) = \lim_{\epsilon\to 0} u_{\mathrm{comp}}(t) = \begin{cases} x_1 \mathrm{csch}(t_1)\sinh(t), & 0 < t < t_1, \\ \infty, & t = 0, \\ \infty, & t = t_1, \end{cases} \tag{3.181}$$

with $u_{\mathrm{comp}}(t)$ given by Eq. (3.169). To write that in a more enlightening form, the Dirac delta function $\delta(t)$ is introduced. The Dirac delta function is defined by its properties

$$\delta(t) = \begin{cases} \infty, & t = 0, \\ 0, & t \neq 0, \end{cases} \tag{3.182}$$

and

$$\int_{-\infty}^{\infty} \delta(t)\,dt = 1. \tag{3.183}$$

The function

$$g_\epsilon(t) = \frac{e^{-\frac{|t|}{\epsilon}}}{2\epsilon} \tag{3.184}$$

is a representation of the Dirac delta function in the limit $\epsilon \to 0$, i.e.,

$$\lim_{\epsilon\to 0} g_\epsilon(t) = \delta(t). \tag{3.185}$$

This can be seen as follows. First, computing the integral yields

$$\int_{-\infty}^{\infty} dt g_\epsilon(t) = 1, \tag{3.186}$$

independent of the value of ϵ. Second, computing the limit gives

$$\lim_{\epsilon\to 0} g_\epsilon(t) = \begin{cases} \lim_{\epsilon\to 0} \frac{1}{2\epsilon} = \infty, & t = 0, \\ \lim_{\epsilon\to 0} \frac{e^{-\frac{|t|}{\epsilon}}}{2\epsilon} = 0, & t \neq 0. \end{cases} \tag{3.187}$$

Noting that $0 \leq t \leq t_1$ and therefore $g_\epsilon(t) = \frac{e^{-\frac{|t|}{\epsilon}}}{2\epsilon} = \frac{e^{-\frac{t}{\epsilon}}}{2\epsilon}$, the control signal can be written in terms of the Dirac delta functions as

$$\begin{aligned} \lim_{\epsilon\to 0} u_{\mathrm{comp}}(t) = {} & x_1 \mathrm{csch}(t_1)\sinh(t) + 2x_1 \mathrm{csch}(t_1)\,\delta(t) \\ & + 2(y_1 - x_1 \coth(t_1))\,\delta(t - t_1). \end{aligned} \tag{3.188}$$

The interpretation is as follows. For $\epsilon = 0$, the boundary layers of the composite solution $y_{\text{comp}}(t)$ degenerate to jumps located exactly at the beginning and end of the time interval. At $t = 0$, the jump leads from the initial condition $y_0 = 0$ to the initial value $y_O(0) = x_1 \operatorname{csch}(t_1)$ of the outer solution. Similarly, at $t = t_1$, the jump leads from the terminal value of the outer solution $y_O(t_1) = x_1 \coth(t_1)$ to the terminal condition y_1. Correspondingly, at the initial time $t = 0$, the control signal is a delta-like impulse which kicks the state component y from its initial value y_0 to $y_O(0)$. The strength of the kick, given by the coefficient $2x_1 \operatorname{csch}(t_1)$ of the Dirac delta function, is two times the jump height of $y(0)$. Intuitively, the reason is that the delta kick is located right at the time domain boundary, and regarding the Dirac delta function as a symmetric function, only half of the kick contributes to the time evolution. Thus, the strength of the kick must be twice as large. Analogously, at the terminal time, a reverse kick occurs with strength $2(y_1 - x_1 \coth(t_1))$, which is twice the height of the jump between $y_O(t_1)$ and the terminal state y_1. This picture remains valid for more complicated unregularized optimal control problems.

3.6. Conclusions

Unregularized optimal trajectory tracking, defined by the target functional Eq. (3.24) with $\epsilon = 0$, is of special interest. Its solution for the controlled state trajectory $\boldsymbol{x}(t)$ can be seen as the limit of realizability for a chosen desired trajectory $\boldsymbol{x}_d(t)$. No other control, be it open or closed loop control, can enforce a state trajectory $\boldsymbol{x}(t)$ with a smaller distance to the desired state trajectory $\boldsymbol{x}_d(t)$. Unregularized trajectory tracking results in singular optimal control. Besides the usual necessary optimality conditions, additional necessary optimality conditions, called the Kelly- or generalized Legendre-Clebsch conditions, have to be satisfied in this case.

The concept of an exactly realizable trajectory, introduced in Chapter 2, is closely related to an unregularized optimal control problem. In fact, if the desired trajectory $\boldsymbol{x}_d(t)$ complies with the initial and terminal conditions

$$\boldsymbol{x}_d(t_0) = \boldsymbol{x}_0, \qquad \boldsymbol{x}_d(t_1) = \boldsymbol{x}_1, \tag{3.189}$$

and satisfies the constraint equation

$$\boldsymbol{\mathcal{Q}}(\boldsymbol{x}_d(t))\left(\dot{\boldsymbol{x}}_d(t) - \boldsymbol{R}(\boldsymbol{x}_d(t))\right) = \boldsymbol{0}, \tag{3.190}$$

then the unregularized optimal control problem of minimizing

$$\begin{aligned}\mathcal{J}[\boldsymbol{x}(t), \boldsymbol{u}(t)] =& \frac{1}{2}\int_{t_0}^{t_1} dt\,(\boldsymbol{x}(t) - \boldsymbol{x}_d(t))^T \boldsymbol{\mathcal{S}}(\boldsymbol{x}(t) - \boldsymbol{x}_d(t)) \\ &+ \frac{1}{2}(\boldsymbol{x}(t_1) - \boldsymbol{x}_1)\boldsymbol{\mathcal{S}}_1(\boldsymbol{x}(t_1) - \boldsymbol{x}_1)\end{aligned} \tag{3.191}$$

subject to

$$\dot{\boldsymbol{x}}(t) = \boldsymbol{R}(\boldsymbol{x}(t)) + \boldsymbol{\mathcal{B}}(\boldsymbol{x}(t))\,\boldsymbol{u}(t), \qquad \boldsymbol{x}(t_0) = \boldsymbol{x}_0 \tag{3.192}$$

is solved by

$$\boldsymbol{x}(t) = \boldsymbol{x}_d(t), \qquad \boldsymbol{u}(t) = \boldsymbol{\mathcal{B}}^{+}(\boldsymbol{x}_d(t))\,(\dot{\boldsymbol{x}}_d(t) - \boldsymbol{R}(\boldsymbol{x}_d(t))). \tag{3.193}$$

The corresponding co-state $\boldsymbol{\lambda}(t)$ vanishes for all times, $\boldsymbol{\lambda}(t) \equiv \boldsymbol{0}$, and the functional $\mathcal{J}$ attains its minimally possible value

$$\mathcal{J}[\boldsymbol{x}(t), \boldsymbol{u}(t)] = 0. \tag{3.194}$$

Furthermore, the controlled state trajectory $\boldsymbol{x}(t)$ and the control signal $\boldsymbol{u}(t)$ are independent of the symmetric matrices of weighting coefficients $\boldsymbol{\mathcal{S}}$ and $\boldsymbol{\mathcal{S}}_1$.

If the linearizing assumption

$$\boldsymbol{\mathcal{Q}}\boldsymbol{R}(\boldsymbol{x}) = \boldsymbol{\mathcal{Q}}\boldsymbol{\mathcal{A}}\boldsymbol{x} + \boldsymbol{\mathcal{Q}}\boldsymbol{b} \tag{3.195}$$

holds, the linear constraint equation (3.190) is readily solved. This results in an exact solution to a nonlinear optimization problem in terms of linear equations. A linear structure underlying nonlinear unregularized optimal trajectory tracking is uncovered. Here, this linear structure is restricted to exactly realizable desired trajectories. This restriction is dropped in the next chapter, which demonstrates that the same underlying linear structure can be exploited for arbitrary desired trajectories and small regularization parameter ϵ.

The exactly solvable linear example in Section 3.5 highlights the difficulties encountered for small ϵ for a desired trajectory which is not exactly realizable. Applying the limit $\epsilon \to 0$ to the exact solution yields an approximation, called the outer solution, which is not uniformly valid over the entire time interval $t_0 \leq t \leq t_1$. Phenomenologically, this non-uniformity manifests in steep transition regions of width ϵ displayed by the state component y close to the initial $t \gtrsim t_0$ and terminal $t \lesssim t_1$ time. These transition regions are known as boundary layers. In analytical approximation, they are described by the inner solutions obtained by rescaling time with ϵ in the exact solution and subsequently applying the limit $\epsilon \to 0$. The matching conditions relate inner and outer solutions in form of overlaps. All inner and outer solutions together with their overlaps are additively combined in a composite solution resulting in a uniformly valid approximation. This procedure is known as a singular perturbation expansion (Bender and Orszag, 2010).

The analytical approximations obtained by singular perturbation expansion reveal a dramatically different behavior of the solution for $\epsilon > 0$ and $\epsilon = 0$. While all state and co-state components are continuous for $\epsilon > 0$, the state component y becomes discontinuous for $\epsilon = 0$. The smooth boundary layers displayed by y degenerate to jumps situated at the time domain boundaries. Even worse, the

control signal, being given by $u(t) = \dot{y}(t)$, diverges at exactly these instants at which $y(t)$ becomes discontinuous. Analytically, these divergences attain the form of Dirac delta functions located at the beginning and end of the time interval. The strength of the delta kicks is twice the height of the corresponding jumps of $y(t)$.

Regarding the behavior of the exact solution with respect to different terminal conditions, the target functional

$$\mathcal{J}[\boldsymbol{x}(t), u(t)] = \frac{1}{2}\int_{t_0}^{t_1}\left((x(t))^2 + (y(t))^2\right)dt + \frac{\epsilon^2}{2}\int_{t_0}^{t_1} dt\,(u(t))^2 + \frac{\beta_1}{2}(x(t_1) - x_1)^2 \tag{3.196}$$

subject to the dynamics constraints

$$\dot{x}(t) = y(t), \quad \dot{y}(t) = u(t), \quad x(0) = x_0, \quad y(0) = y_0, \quad y(t_1) = y_1, \tag{3.197}$$

constitutes the most general one in the limit of small ϵ. All variants of sharp and free terminal conditions can be generated from the solution to Eq. (3.196) by applying the limits $\beta_1 \to \infty$ and $\beta_1 \to 0$. A terminal term of the form $\beta_2 (y(t_1) - y_1)^2$ in Eq. (3.196) becomes irrelevant as $\epsilon \to 0$, and the solution becomes independent of β_2. The limits $\beta_1 \to \infty$ and $\beta_1 \to 0$ commute with the determination of the leading order approximation for small ϵ.

4. Analytical approximations for optimal trajectory tracking

Chapters 2 and 3 uncovered an underlying linear structure of unregularized nonlinear optimal trajectory tracking for exactly realizable desired trajectories $\boldsymbol{x}_d(t)$. Clearly, the assumption of $\boldsymbol{x}_d(t)$ to be exactly realizable is overly restrictive. Only p components of $\boldsymbol{x}_d(t)$ can be prescribed by the experimenter, while the remaining $n-p$ components are fixed by the constraint equation. This approach seriously limits the true power of optimal control, which guarantees the existence of solutions for a huge class of desired trajectories.

This chapter drops the assumption of exactly realizable trajectories and allows for arbitrary desired trajectories $\boldsymbol{x}_d(t)$. The regularization parameter ϵ is assumed to be small, $\epsilon \ll 1$, and used for a perturbation expansion. Rearranging the necessary optimality condition leads to a reinterpretation of unregularized optimal control problems as singularly perturbed differential equations. For systems satisfying a linearizing assumption, the leading order equations become linear. The linearity allows the derivation of closed form expressions for optimal trajectory tracking in a general class of nonlinear systems affine in control. The perturbative approach yields exact results for $\epsilon = 0$. However, this exact result comes at a price. The limit $\epsilon \to 0$ leads to a diverging control signal and a discontinuous state trajectory.

The analytical approach applies to mechanical control systems defined in Example 1.1 as well as to the activator-controlled FHN model of Example 1.2. Section 4.1 presents the straightforward but tedious derivation for a general model comprising both examples. A comparison with numerical solutions of optimal control is performed in Section 4.2. Analytical solutions to optimal control problems also apply to optimal feedback control. Continuous-time and time-delayed feedback are discussed in Section 4.3. Section 4.4 tackles the singular perturbation expansion of general dynamical systems, and Section 4.5 draws conclusions.

4.1. Two-dimensional dynamical systems

4.1.1. General procedure

The analytic approach to optimal trajectory tracking is a straightforward application of singular perturbation theory. The first step consists in rearranging the necessary

optimality conditions in Section 4.1.3. This rearrangement allows the interpretation of a singular optimal control problem as a singularly perturbed system of differential equations. The small regularization parameter ϵ multiplies the highest order derivative of the system. Setting $\epsilon = 0$ changes the differential order of the system and results in a violation of initial conditions. The equations thus obtained are called the *outer equations*. Their solution is discussed in Section 4.1.4. The outer solutions are not uniformly valid over the whole time interval. The situation is analogous to the exactly solvable example from Section 3.5. Similar measures are taken to resolve the problem. The time is rescaled by ϵ, and a different set of equations called the *inner equations* is derived. Their solutions are able to accommodate all initial and terminal conditions. The inner equations are valid close to the initial and terminal conditions. Eventually, both sets of solutions have to be connected by the *matching procedure*. Several free constants of inner and outer solutions must be determined by matching conditions. The inner equations, their solutions as well as matching is discussed in Section 4.1.5. It remains to combine all inner and outer solutions in a single *composite solution.* Only the composite solution yields a uniformly valid approximation over the whole time domain. The control signal is given in terms of the composite solution. Both points are discussed in Section 4.1.6. The last step involves a discussion of the exact solution obtained for $\epsilon = 0$ in Section 4.1.7.

4.1.2. Necessary optimality conditions

The solution to optimal trajectory tracking is derived for nonlinear two-dimensional dynamical systems of the form

$$\dot{x}(t) = a_0 + a_1 x(t) + a_2 y(t), \tag{4.1}$$

$$\dot{y}(t) = R(x(t), y(t)) + b(x(t), y(t))\, u(t). \tag{4.2}$$

The parameters a_0 and a_1 are arbitrary. The function $b(x, y)$ is not allowed to vanish for any value of x and y. The system is controllable as long as $a_2 \neq 0$. This was proven in Example 2.11 of Section 2.4.3, and $a_2 \neq 0$ is assumed in the following. The optimal control problem is the minimization of the functional

$$\begin{aligned}\mathcal{J}[\boldsymbol{x}(t), u(t)] = \frac{1}{2}\int_{t_0}^{t_1} \left(s_1 (x(t) - x_d(t))^2 + s_2 (y(t) - y_d(t))^2\right) dt \\ + \frac{\beta_1}{2}(x(t_1) - x_1)^2 + \frac{\epsilon^2}{2}\int_{t_0}^{t_1} dt\, (u(t) - u_0)^2 .\end{aligned} \tag{4.3}$$

The minimization of Eq. (4.3) is constrained by the condition that $x(t)$ and $y(t)$ are a solution of the controlled dynamical system Eqs. (4.1), (4.2), together with the initial and terminal conditions

$$x(t_0) = x_0, \qquad y(t_0) = y_0, \qquad y(t_1) = y_1. \tag{4.4}$$

In contrast to $x(t)$, the y-component satisfies a sharp terminal condition. This special choice for the terminal conditions is motivated by the exact solution discussed in Section 3.5. All relevant cases of terminal conditions are covered by the weighting coefficient $0 \leq \beta_1 \leq \infty$. The constants $s_1 > 0$ and $s_2 > 0$ are positive weights which correspond to a positive definite diagonal matrix of weighting coefficients,

$$\boldsymbol{\mathcal{S}} = \begin{pmatrix} s_1 & 0 \\ 0 & s_2 \end{pmatrix}. \tag{4.5}$$

Equation (4.3) takes a constant background value u_0 into account. For a vanishing regularization coefficient $\epsilon = 0$, the minimization problem becomes singular. The perturbation expansion applies in the limit of small $\epsilon > 0$. The co-state is denoted by

$$\boldsymbol{\lambda}(t) = \begin{pmatrix} \lambda_x(t) \\ \lambda_y(t) \end{pmatrix}. \tag{4.6}$$

The adjoint or co-state equation involves the Jacobian $\nabla \boldsymbol{R}$ of the nonlinearity $\boldsymbol{R}$,

$$\nabla \boldsymbol{R}(\boldsymbol{x}(t)) = \begin{pmatrix} a_1 & a_2 \\ \partial_x R(x(t), y(t)) & \partial_y R(x(t), y(t)) \end{pmatrix}. \tag{4.7}$$

The necessary optimality conditions are (see Section 3.1 for a discussion of the solution to general linear dynamical systems)

$$0 = \epsilon^2 (u(t) - u_0) + b(x(t), y(t)) \lambda_y(t), \tag{4.8}$$

$$\begin{pmatrix} \dot{x}(t) \\ \dot{y}(t) \end{pmatrix} = \begin{pmatrix} y(t) \\ R(x(t), y(t)) \end{pmatrix} + b(x(t), y(t)) \begin{pmatrix} 0 \\ u(t) \end{pmatrix}, \tag{4.9}$$

$$\begin{aligned} -\begin{pmatrix} \dot{\lambda}_x(t) \\ \dot{\lambda}_y(t) \end{pmatrix} &= \begin{pmatrix} a_1 & \partial_x R(x(t), y(t)) + \partial_x b(x(t), y(t)) u(t) \\ a_2 & \partial_y R(x(t), y(t)) + \partial_y b(x(t), y(t)) u(t) \end{pmatrix} \begin{pmatrix} \lambda_x(t) \\ \lambda_y(t) \end{pmatrix} \\ &\quad + \begin{pmatrix} s_1 (x(t) - x_d(t)) \\ s_2 (y(t) - y_d(t)) \end{pmatrix}, \end{aligned} \tag{4.10}$$

$$\begin{pmatrix} x(t_0) \\ y(t_0) \end{pmatrix} = \begin{pmatrix} x_0 \\ y_0 \end{pmatrix}, \tag{4.11}$$

$$y(t_1) = y_1, \tag{4.12}$$

$$\lambda_x(t_1) = \beta_1 (x(t_1) - x_1). \tag{4.13}$$

4.1.3. Rearranging the necessary optimality conditions

The first step is to solve Eq. (4.9) for the control signal,

$$u(t) = \frac{1}{b(x(t), y(t))} (\dot{y}(t) - R(x(t), y(t))). \tag{4.14}$$

Using $u(t)$ in Eq. (4.8) yields

$$\lambda_y(t) = -\frac{\epsilon^2}{b(x(t), y(t))^2}\left(\dot{y}(t) - R(x(t), y(t)) - b(x(t), y(t))\, u_0\right). \tag{4.15}$$

This expression holds for all times t such that it is allowed to apply the time derivative,

$$\begin{aligned}\dot{\lambda}_y(t) = &-\epsilon^2 \frac{1}{b(x(t), y(t))^2} \dot{x}(t) \left(u_0 \partial_x b(x(t), y(t)) - \partial_x R(x(t), y(t))\right) \\ &- \epsilon^2 \frac{1}{b(x(t), y(t))^2} \left(\dot{y}(t) \left(u_0 \partial_y b(x(t), y(t)) - \partial_y R(x(t), y(t))\right) + \ddot{y}(t)\right) \\ &- \epsilon^2 \frac{2\left(R(x(t), y(t)) - \dot{y}(t)\right)}{b(x(t), y(t))^3} \left(\dot{x}(t)\, \partial_x b(x(t), y(t)) + \dot{y}(t)\, \partial_y b(x(t), y(t))\right).\end{aligned} \tag{4.16}$$

Using both relations Eqs. (4.15) and Eq. (4.16), all occurrences of $\lambda_y(t)$ can be eliminated in the co-state equations (4.10). The equation for λ_x becomes

$$\begin{aligned}-\dot{\lambda}_x(t) = &\, a_1 \lambda_x(t) + s_1 (x(t) - x_d(t)) \\ &- \frac{\epsilon^2}{b(x(t), y(t))^2} \left(\partial_x R(x(t), y(t)) + \partial_x b(x(t), y(t))\, u(t)\right) \\ &\times \left(\dot{y}(t) - R(x(t), y(t)) - b(x(t), y(t))\, u_0\right),\end{aligned} \tag{4.17}$$

while the equation for λ_y transforms to a second order differential equation for y,

$$\begin{aligned}&\epsilon^2 \ddot{y}(t) + \epsilon^2 \dot{x}(t) \left(u_0 \partial_x b(x(t), y(t)) - \partial_x R(x(t), y(t))\right) \\ &+ \epsilon^2 \dot{y}(t) \left(u_0 \partial_y b(x(t), y(t)) - \partial_y R(x(t), y(t))\right) \\ &+ 2\epsilon^2 \frac{\left(R(x(t), y(t)) - \dot{y}(t)\right)}{b(x(t), y(t))} w_1(x(t), y(t)) \\ &= b(x(t), y(t))^2 \left(a_2 \lambda_x(t) + s_2 (y(t) - y_d(t))\right) \\ &+ \epsilon^2 w_2(x(t), y(t)) \left(b(x(t), y(t))\, u_0 - \left(\dot{y}(t) - R(x(t), y(t))\right)\right).\end{aligned} \tag{4.18}$$

Here, w_1 and w_2 denote the abbreviations

$$w_1(x(t), y(t)) = \dot{x}(t)\, \partial_x b(x(t), y(t)) + \dot{y}(t)\, \partial_y b(x(t), y(t)), \tag{4.19}$$

$$w_2(x(t), y(t)) = \partial_y R(x(t), y(t)) + \frac{\partial_y b(x(t), y(t))}{b(x(t), y(t))} \left(\dot{y}(t) - R(x(t), y(t))\right). \tag{4.20}$$

The equation for $x(t)$ does not change,

$$\dot{x}(t) = a_0 + a_1 x(t) + a_2 y(t). \tag{4.21}$$

The system of equations (4.17)-(4.21) must be solved with four initial and terminal conditions Eqs. (4.11)-(4.13). The rearranged equations look horrible and much more difficult than before. However, the small parameter ϵ^2 multiplies the second order derivative of $y(t)$ as well as every occurrence of the nonlinear force term R and its derivatives. Due to ϵ multiplying the highest order derivative, Eqs. (4.17)-(4.21) constitute a singularly perturbed system of differential equations. Setting $\epsilon = 0$ changes the order of the system. Consequently, not all four boundary conditions Eqs. (4.11)-(4.13) can be satisfied. Singular perturbation theory has to be applied to solve Eqs. (4.17)-(4.21) in the limit $\epsilon \to 0$.

4.1.4. Outer equations

The outer equations are defined for the ordinary time scale t. The outer solutions are denoted with index O,

$$x_O(t) = x(t), \qquad y_O(t) = y(t), \qquad \lambda_O(t) = \lambda_x(t). \tag{4.22}$$

Expanding Eqs. (4.17)-(4.21) up to leading order in ϵ yields two linear differential equations of first order and an algebraic equation,

$$\dot{\lambda}_O(t) = -a_1\lambda_x(t) + s_1(x_d(t) - x_O(t)), \tag{4.23}$$

$$\lambda_O(t) = \frac{s_2}{a_2}(y_d(t) - y_O(t)), \tag{4.24}$$

$$\dot{x}_O(t) = a_0 + a_1 x_O(t) + a_2 y_O(t). \tag{4.25}$$

Equations (4.23)-(4.25) allow for two initial conditions,

$$x_O(t_0) = x_{\text{init}}, \qquad y_O(t_0) = y_{\text{init}}. \tag{4.26}$$

Note that x_{init} and y_{init} are not given by the initial conditions (4.11) but have to be determined by matching with the inner solutions. Eliminating $\lambda_x(t)$ from Eqs. (4.23)-(4.25) yields two coupled ODEs for $x_O(t)$ and $y_O(t)$,

$$\dot{x}_O(t) = a_1 x_O(t) + a_2 y_O(t) + a_0, \tag{4.27}$$

$$\dot{y}_O(t) = -a_1 y_O(t) + \frac{a_2 s_1}{s_2} x_O(t) + \dot{y}_d(t) + a_1 y_d(t) - \frac{a_2 s_1}{s_2} x_d(t). \tag{4.28}$$

It is convenient to express the solutions for x_O and y_O in terms of the state transition matrix $\mathbf{\Phi}(t, t_0)$ (see Appendix A.1 for a general derivation of state transition matrices)

$$\begin{pmatrix} x_O(t) \\ y_O(t) \end{pmatrix} = \mathbf{\Phi}(t, t_0) \begin{pmatrix} x_{\text{init}} \\ y_{\text{init}} \end{pmatrix} + \int_{t_0}^{t} d\tau \mathbf{\Phi}(t, \tau) \boldsymbol{f}(\tau), \tag{4.29}$$

with

$$\boldsymbol{\Phi}\left(t,t_0\right)=\begin{pmatrix}\cosh\left(\Delta t\varphi_1\right)+\frac{a_1}{\varphi_1}\sinh\left(\Delta t\varphi_1\right) & \frac{a_2}{\varphi_1}\sinh\left(\Delta t\varphi_1\right)\\ \frac{a_2 s_1}{s_2\varphi_1}\sinh\left(\Delta t\varphi_1\right) & \cosh\left(\Delta t\varphi_1\right)-\frac{a_1}{\varphi_1}\sinh\left(\Delta t\varphi_1\right)\end{pmatrix}, \tag{4.30}$$

$$\varphi_1=\frac{\sqrt{a_1^2 s_2+a_2^2 s_1}}{\sqrt{s_2}},\ \Delta t=t-t_0, \tag{4.31}$$

and inhomogeneity

$$\boldsymbol{f}\left(t\right)=\begin{pmatrix}a_0\\ \dot{y}_d\left(t\right)+a_1 y_d\left(t\right)-\frac{a_2 s_1}{s_2}x_d\left(t\right)\end{pmatrix}. \tag{4.32}$$

For later reference, the solutions for $x_O\left(t\right)$ and $y_O\left(t\right)$ are given explicitly,

$$\begin{aligned}x_O\left(t\right)=&\frac{a_1 a_2}{\varphi_1}\int_{t_0}^{t} y_d\left(\tau\right)\sinh\left(\varphi_1\left(t-\tau\right)\right)\,d\tau+a_2\int_{t_0}^{t} y_d\left(\tau\right)\cosh\left(\varphi_1\left(t-\tau\right)\right)\,d\tau\\ &-\frac{a_2^2 s_1}{s_2\varphi_1}\int_{t_0}^{t} x_d\left(\tau\right)\sinh\left(\varphi_1\left(t-\tau\right)\right)\,d\tau+\frac{1}{\varphi_1}\sinh\left(\left(t-t_0\right)\varphi_1\right)\left(a_0-a_2 y_d\left(t_0\right)\right)\\ &+\frac{a_0 a_1}{\varphi_1^2}\left(\cosh\left(\left(t-t_0\right)\varphi_1\right)-1\right)\\ &+x_{\text{init}}\left(\frac{a_1}{\varphi_1}\sinh\left(\left(t-t_0\right)\varphi_1\right)+\cosh\left(\left(t-t_0\right)\varphi_1\right)\right)\\ &+\frac{a_2}{\varphi_1}y_{\text{init}}\sinh\left(\left(t-t_0\right)\varphi_1\right),\end{aligned} \tag{4.33}$$

$$\begin{aligned}y_O\left(t\right)=&\frac{a_0 a_2 s_1}{s_2\varphi_1^2}\left(\cosh\left(\left(t-t_0\right)\varphi_1\right)-1\right)+\frac{a_2 s_1}{s_2\varphi_1}x_{\text{init}}\sinh\left(\left(t-t_0\right)\varphi_1\right)\\ &+y_{\text{init}}\cosh\left(\left(t-t_0\right)\varphi_1\right)-y_{\text{init}}\frac{a_1}{\varphi_1}\sinh\left(\left(t-t_0\right)\varphi_1\right)\\ &-\frac{a_2 s_1}{s_2}\int_{t_0}^{t} x_d\left(\tau\right)\cosh\left(\varphi_1\left(t-\tau\right)\right)\,d\tau+\frac{a_2 s_1 a_1}{s_2\varphi_1}\int_{t_0}^{t} x_d\left(\tau\right)\sinh\left(\varphi_1\left(t-\tau\right)\right)\,d\tau\\ &+\frac{a_2^2 s_1}{s_2\varphi_1}\int_{t_0}^{t} y_d\left(\tau\right)\sinh\left(\varphi_1\left(t-\tau\right)\right)\,d\tau\\ &+\frac{a_1}{\varphi_1}y_d\left(t_0\right)\sinh\left(\left(t-t_0\right)\varphi_1\right)+y_d\left(t\right)-y_d\left(t_0\right)\cosh\left(\left(t-t_0\right)\varphi_1\right).\end{aligned} \tag{4.34}$$

The solution for the co-state $\lambda_O\left(t\right)$ reads

$$\lambda_O\left(t\right)=\frac{s_2}{a_2}\left(y_d\left(t\right)-y_O\left(t\right)\right). \tag{4.35}$$

Evaluating Eq. (4.35) at the initial and terminal time t_0 and t_1 yields relations for $\lambda_O(t_0)$ and $\lambda_O(t_1)$, respectively,

$$\lambda_O(t_0) = \frac{s_2}{a_2}\left(y_d(t_0) - y_{\text{init}}\right), \tag{4.36}$$

$$\lambda_O(t_1) = \frac{s_2}{a_2}\left(y_d(t_1) - y_{\text{end}}\right). \tag{4.37}$$

The abbreviation y_{end} is

$$y_{\text{end}} = y_O(t_1). \tag{4.38}$$

Equations (4.36)-(4.38) will be useful for matching.

4.1.5. Inner equations

4.1.5.1. Initial boundary layer

Boundary layers occur at both ends of the time domain. The initial boundary layer at the left end of the time domain is resolved using the time scale $\tau_L = (t - t_0)/\epsilon$ and scaled solutions

$$X_L(\tau_L) = X_L\left((t - t_0)/\epsilon\right) = x(t) = x(t_0 + \epsilon\tau_L), \tag{4.39}$$

$$Y_L(\tau_L) = Y_L\left((t - t_0)/\epsilon\right) = y(t) = y(t_0 + \epsilon\tau_L), \tag{4.40}$$

$$\Lambda_L(\tau_L) = \Lambda_L\left((t - t_0)/\epsilon\right) = \lambda_x(t) = \lambda_x(t_0 + \epsilon\tau_L). \tag{4.41}$$

From the definitions of X_L, Y_L, and Λ_L together with the initial conditions for x and y, Eqs. (4.11), follow the initial conditions

$$x(t_0) = X_L(0) = x_0, \qquad y(t_0) = Y_L(0) = y_0. \tag{4.42}$$

The derivatives of x transform as

$$\dot{x}(t) = \frac{1}{\epsilon}X_L'(\tau_L), \qquad \ddot{x}(t) = \frac{1}{\epsilon^2}X_L''(\tau_L), \tag{4.43}$$

and analogously for y and λ_x. The prime $X_L'(\tau_L)$ denotes the derivative of X_L with respect to its argument. The matching conditions at the left boundary layer are

$$\lim_{t\to t_0} x_O(t) = \lim_{\tau_L\to\infty} X_L(\tau_L), \tag{4.44}$$

$$\lim_{t\to t_0} y_O(t) = \lim_{\tau_L\to\infty} Y_L(\tau_L), \tag{4.45}$$

$$\lim_{t\to t_0} \lambda_O(t) = \lim_{\tau_L\to\infty} \Lambda_L(\tau_L). \tag{4.46}$$

Using the definitions Eqs. (4.39)-(4.41) in Eqs. (4.17)-(4.21) and expanding in ϵ yields the left inner equations in leading order as

$$\Lambda_L'(\tau_L) = 0, \tag{4.47}$$

$$\begin{aligned} Y_L''(\tau_L) = {} & Y_L'(\tau_L)^2 \frac{\partial_y b(X_L(\tau_L), Y_L(\tau_L))}{b(X_L(\tau_L), Y_L(\tau_L))} + 2X_L'(\tau_L) Y_L'(\tau_L) \frac{\partial_x b(X_L(\tau_L), Y_L(\tau_L))}{b(X_L(\tau_L), Y_L(\tau_L))} \\ & + b(X_L(\tau_L), Y_L(\tau_L))^2 (s_2 (Y_L(\tau_L) - y_d(t_0)) + a_2 \Lambda_L(\tau_L)), \end{aligned} \tag{4.48}$$

$$X_L'(\tau_L) = 0. \tag{4.49}$$

The differential equations do not involve the nonlinearity R. The solutions for Λ_L and X_L are

$$\Lambda_L(\tau_L) = \Lambda_{L,0} = \lambda_O(t_0) = \frac{s_2}{a_2}(y_d(t_0) - y_{\text{init}}), \tag{4.50}$$

$$X_L(\tau_L) = x_0. \tag{4.51}$$

To obtain the value for $\Lambda_{L,0}$, Eq. (4.36) was used together with the matching condition Eq. (4.46). The matching condition for x, Eq. (4.44), immediately yields

$$x_{\text{init}} = x_0, \tag{4.52}$$

while y_{init} will be determined later on. The equation for Y_L simplifies to

$$Y_L''(\tau_L) = Y_L'(\tau_L)^2 \frac{\partial_y b(x_0, Y_L(\tau_L))}{b(x_0, Y_L(\tau_L))} + s_2 b(x_0, Y_L(\tau_L))^2 (Y_L(\tau_L) - y_{\text{init}}). \tag{4.53}$$

As long as b depends on Y_L, this is a nonlinear equation. Because it is autonomous, it can be transformed to a first order ODE by introducing a new function v_L defined as

$$Y_L'(\tau_L) = v_L(Y_L(\tau_L)) b(x_0, Y_L(\tau_L)) \tag{4.54}$$

such that the second order time derivative is

$$\begin{aligned} Y_L''(\tau_L) = {} & v_L'(Y_L(\tau_L)) v_L(Y_L(\tau_L)) b(x_0, Y_L(\tau_L))^2 \\ & + v_L(Y_L(\tau_L))^2 \partial_y b(x_0, Y_L(\tau_L)) b(x_0, Y_L(\tau_L)). \end{aligned} \tag{4.55}$$

This leads to a fairly simple equation for v_L,

$$\frac{1}{2} \partial_{Y_L} (v_L(Y_L))^2 = s_2 (Y_L - y_{\text{init}}). \tag{4.56}$$

Equation (4.56) is to be solved with the matching condition Eq. (4.45),

$$\lim_{\tau_L \to \infty} Y_L(\tau_L) = \lim_{t \to t_0} y_O(t) = y_{\text{init}}. \tag{4.57}$$

A limit $\lim_{\tau_L \to \infty} Y_L(\tau_L)$ exists if $Y_L(\tau_L)$ is neither infinite nor oscillatory as τ_L approaches ∞. The existence of both $\lim_{\tau_L \to \infty} Y_L(\tau_L)$ and $\lim_{\tau_L \to \infty} Y_L'(\tau_L)$ implies

$$\lim_{\tau_L \to \infty} Y_L'(\tau_L) = 0. \tag{4.58}$$

A proof is straightforward. From Eq. (4.57) follows

$$\lim_{\tau_L \to \infty} \frac{Y_L(\tau_L)}{\tau_L} = \lim_{\tau_L \to \infty} \frac{y_{\text{init}}}{\tau_L} = 0, \tag{4.59}$$

and applying L'Hôpital's rule to

$$\lim_{\tau_L \to \infty} \frac{Y_L(\tau_L) - y_{\text{init}}}{\tau_L} = \lim_{\tau_L \to \infty} \frac{Y_L(\tau_L)}{\tau_L} - \lim_{\tau_L \to \infty} \frac{y_{\text{init}}}{\tau_L} = 0, \tag{4.60}$$

and

$$\lim_{\tau_L \to \infty} \frac{Y_L(\tau_L) - y_{\text{init}}}{\tau_L} = \frac{0}{0} = \lim_{\tau_L \to \infty} Y_L'(\tau_L) = 0 \tag{4.61}$$

yields the result.

From Eq. (4.58) follows the initial condition for v_L as

$$v_L(y_{\text{init}}) = 0. \tag{4.62}$$

Solving Eq. (4.56) with this condition yields

$$v_L(Y_L) = \pm\sqrt{s_2}\,|y_{\text{init}} - Y_L|\,. \tag{4.63}$$

The solution Eq. (4.63) together with the definition for v_L, Eq. (4.54), leads to a first order nonlinear ODE for Y_L,

$$Y_L'(\tau_L) = \pm\sqrt{s_2}\,|y_{\text{init}} - Y_L(\tau_L)|\,b(x_0, Y_L(\tau_L))\,. \tag{4.64}$$

The last point is to determine which sign in Eq. (4.64) is the relevant one. Note that $Y_L(\tau_L) = y_{\text{init}}$ is a stationary point of Eq. (4.64) which cannot be crossed by the dynamics. Furthermore, this is the only stationary point. Because of $b(x_0, Y_L(\tau_L)) \neq 0$ by assumption, b cannot change its sign. If initially $y_{\text{init}} > Y_L(0) = y_0$ and $b(x_0, Y_L(\tau_L)) > 0$ for all times $\tau_L > 0$, Y_L must grow and therefore $Y_L' = \sqrt{s_2}\,(y_{\text{init}} - Y_L)\,b(x_0, Y_L) > 0$ is the correct choice. On the other hand, if $y_{\text{init}} > Y_L(0) = y_0$ and $b(x_0, Y_L(\tau_L)) < 0$ for all times $\tau_L > 0$, Y_L must decrease and consequently Y_L evolves according to $Y_L' = -\sqrt{s_2}\,(y_{\text{init}} - Y_L)\,b(x_0, Y_L) > 0$. These considerations finally lead to

$$Y_L'(\tau_L) = \sqrt{s_2}\,(y_{\text{init}} - Y_L(\tau_L))\,|b(x_0, Y_L(\tau_L))|\,, \tag{4.65}$$

$$Y_L(0) = y_0. \tag{4.66}$$

An analytical solution of Eq. (4.65) for arbitrary functions b does not exist in closed form. If $b(x, y) = b(x)$ does not depend on y, Eq. (4.65) is linear and has the solution

$$Y_L(\tau_L) = y_{\text{init}} + \exp\left(-\sqrt{s_2}\,|b(x_0)|\,\tau_L\right)(y_0 - y_{\text{init}})\,. \tag{4.67}$$

4.1.5.2. Terminal boundary layer

A treatment analogous to Section 4.1.5.1 is performed to resolve the boundary layer at the right end of the time domain. The relevant time scale is $\tau_R = (t_1 - t)/\epsilon$, and the scaled solutions are defined as

$$X_R(\tau_R) = X_R((t_1 - t)/\epsilon) = x(t), \tag{4.68}$$
$$Y_R(\tau_R) = Y_R((t_1 - t)/\epsilon) = y(t), \tag{4.69}$$
$$\Lambda_R(\tau_R) = \Lambda_R((t_1 - t)/\epsilon) = \lambda_x(t). \tag{4.70}$$

The terminal conditions Eqs. (4.13) and (4.12) lead to the boundary conditions

$$\Lambda_R(0) = \beta_1 (X_R(0) - x_1), \qquad Y_R(0) = y_1. \tag{4.71}$$

Furthermore, X_R, Y_R, and Λ_R have to satisfy the matching conditions

$$\lim_{t \to t_1} x_O(t) = \lim_{\tau_R \to \infty} X_R(\tau_R), \tag{4.72}$$
$$\lim_{t \to t_1} y_O(t) = \lim_{\tau_R \to \infty} Y_R(\tau_R), \tag{4.73}$$
$$\lim_{t \to t_1} \lambda_O(t) = \lim_{\tau_R \to \infty} \Lambda_R(\tau_R). \tag{4.74}$$

The derivatives of x transform as

$$\dot{x}(t) = -\frac{1}{\epsilon} X_R'(\tau_R), \qquad \ddot{x}(t) = \frac{1}{\epsilon^2} X_R''(\tau_R), \tag{4.75}$$

and analogously for y and λ_x. Plugging these definitions in Eqs. (4.17)-(4.21) and expanding in ϵ yields the right inner equations in leading order,

$$\Lambda_R'(\tau_R) = 0, \tag{4.76}$$
$$Y_R''(\tau_R) = Y_R'(\tau_R)^2 \frac{\partial_y b(X_R(\tau_R), Y_R(\tau_R))}{b(X_R(\tau_R), Y_R(\tau_R))} + 2X_R'(\tau_R) Y_R'(\tau_R) \frac{\partial_x b(X_R(\tau_R), Y_R(\tau_R))}{b(X_R(\tau_R), Y_R(\tau_R))} + b(X_R(\tau_R), Y_R(\tau_R))^2 (s_2 (Y_R(\tau_R) - y_d(t_1)) + a_2 \Lambda_R(\tau_R)), \tag{4.77}$$
$$X_R'(\tau_R) = 0. \tag{4.78}$$

These equations are identical in form to the left inner equations (4.47)-(4.49). The solutions for $X_R(\tau_R)$ and $\Lambda_R(\tau_R)$ are constant and can be written as

$$X_R(\tau_R) = x_1 + \frac{1}{\beta_1} \Lambda_{R,0}, \qquad \Lambda_R(\tau_R) = \Lambda_{R,0}. \tag{4.79}$$

Applying the matching condition Eq. (4.74) together with Eq. (4.37) yields the solution for $\Lambda_R(\tau_R)$ and $X_R(\tau_R)$ as

$$\Lambda_R(\tau_R) = \Lambda_{R,0} = \lambda_O(t_1) = \frac{s_2}{a_2} (y_d(t_1) - y_{\text{end}}), \tag{4.80}$$
$$X_R(\tau_R) = x_1 + \frac{s_2}{\beta_1 a_2} (y_d(t_1) - y_{\text{end}}). \tag{4.81}$$

With the analogous considerations as for the left inner equations, see Eq. (4.65), the solution to $Y_R(\tau_R)$ is given by the first order ODE

$$Y_R'(\tau_R) = \sqrt{s_2}\,(y_{\text{end}} - Y_R(\tau_R))\,|b(x_1, Y_R(\tau_R))|\,, \tag{4.82}$$

$$Y_R(0) = y_1. \tag{4.83}$$

Equation (4.82) satisfies the matching condition Eq. (4.73). The remaining matching condition Eq. (4.72) gives

$$x_O(t_1) = x_1 + \frac{s_2}{\beta_1 a_2}\,(y_d(t_1) - y_{\text{end}})\,. \tag{4.84}$$

The constant $y_{\text{end}} = y_O(t_1)$ depends on y_{init}, which is the last free parameter of the outer solution. Solving Eq. (4.84) for y_{init} yields

$$\begin{aligned}
y_{\text{init}} = {} & \frac{s_2\varphi_1}{\kappa}\sinh((t_0 - t_1)\,\varphi_1)\left(\left(a_1 s_2 - a_2^2\beta_1\right)y_d(t_0) + a_2\left(a_0\beta_1 + x_0\left(a_1\beta_1 + s_1\right)\right)\right) \\
& - \frac{1}{\kappa}\cosh((t_1 - t_0)\,\varphi_1)\left(a_2\left(\beta_1\left(a_1^2 s_2 + a_2^2 s_1\right)x_0 + a_0 s_2\left(a_1\beta_1 + s_1\right)\right)\right) \\
& + \frac{s_2}{\kappa}\left(a_1^2 s_2 + a_2^2 s_1\right)\cosh((t_1 - t_0)\,\varphi_1)\,y_d(t_0) \\
& + \frac{a_2\varphi_1 s_1}{\kappa}\left(a_2^2\beta_1 - a_1 s_2\right)\int_{t_0}^{t_1} x_d(\tau)\sinh(\varphi_1(t_1 - \tau))\;d\tau \\
& - \frac{a_2^2\varphi_1 s_2}{\kappa}\left(a_1\beta_1 + s_1\right)\int_{t_0}^{t_1} y_d(\tau)\sinh(\varphi_1(t_1 - \tau))\;d\tau \\
& - \frac{\beta_1 a_2^2\varphi_1^2 s_2}{\kappa}\int_{t_0}^{t_1} y_d(\tau)\cosh(\varphi_1(t_1 - \tau))\;d\tau + \frac{a_2 a_0 s_2}{\kappa}\left(a_1\beta_1 + s_1\right) \\
& + \frac{a_2\varphi_1^2 s_1 s_2}{\kappa}\int_{t_0}^{t_1} x_d(\tau)\cosh(\varphi_1(t_1 - \tau))\;d\tau + \beta_1 x_1\frac{a_2}{\kappa}\left(a_1^2 s_2 + a_2^2 s_1\right).
\end{aligned} \tag{4.85}$$

Equation (4.85) contains the abbreviation

$$\kappa = s_2\varphi_1\left(a_2^2\beta_1 - a_1 s_2\right)\sinh((t_1 - t_0)\,\varphi_1) + s_2^2\varphi_1^2\cosh((t_1 - t_0)\,\varphi_1)\,. \tag{4.86}$$

Finally, all inner and outer solutions are determined; and all matching conditions are satisfied.

Example 4.1: Consistency check

For appropriate parameter values, the solution derived in this section must reduce to the corresponding leading order approximation for the exact solution of Section 3.5. For simplicity, only the case of vanishing initial conditions, $x_0 = y_0 = 0$, is considered. The other parameters have values

$$a_0 = 0, \qquad a_1 = 0, \qquad a_2 = 1, \qquad s_1 = 1, \qquad s_2 = 1, \qquad t_0 = 0. \tag{4.87}$$

The desired trajectories and the coupling function are

$$x_d(t) \equiv 0, \qquad y_d(t) \equiv 0, \qquad b(x, y) \equiv 1. \tag{4.88}$$

From these assumptions follows

$$\varphi_1 = \frac{\sqrt{a_1^2 s_2 + a_2^2 s_1}}{\sqrt{s_2}} = 1, \quad y_{\text{init}} = \frac{\beta_1 x_1}{\kappa}, \quad \kappa = \beta_1 \sinh(t_1) + \cosh(t_1). \tag{4.89}$$

The outer solution of the controlled state is obtained as

$$\begin{pmatrix} x_O(t) \\ y_O(t) \\ \lambda_O(t) \end{pmatrix} = \begin{pmatrix} \dfrac{\beta_1 x_1}{\kappa} \sinh(t) \\ \dfrac{\beta_1 x_1}{\kappa} \cosh(t) \\ -\dfrac{\beta_1 x_1}{\kappa} \cosh(t) \end{pmatrix}. \tag{4.90}$$

This is indeed the outer limit of the exact solution from Section 3.5, Eq. (3.141) with $x_0 = y_0 = 0$. The abbreviation y_{end} simplifies to

$$y_{\text{end}} = y_O(t_1) = \frac{\beta_1 x_1}{\kappa} \cosh(t_1). \tag{4.91}$$

The solution to the left inner equations yields

$$\begin{pmatrix} X_L(\tau_L) \\ Y_L(\tau_L) \\ \Lambda_L(\tau_L) \end{pmatrix} = \begin{pmatrix} 0 \\ \dfrac{\beta_1 x_1}{\kappa}(1 - e^{-\tau_L}) \\ -\dfrac{\beta_1 x_1}{\kappa} \end{pmatrix}, \tag{4.92}$$

while the solutions to the right inner equations becomes

$$\begin{pmatrix} X_R(\tau_R) \\ Y_R(\tau_R) \\ \Lambda_R(\tau_R) \end{pmatrix} = \begin{pmatrix} \frac{\beta_1 x_1}{\kappa} \sinh(t_1) \\ \frac{\beta_1 x_1}{\kappa} \cosh(t_1) + e^{-\tau_R} \left(y_1 - \frac{\beta_1 x_1}{\kappa} \cosh(t_1) \right) \\ -\frac{\beta_1 x_1}{\kappa} \cosh(t_1) \end{pmatrix}. \tag{4.93}$$

The left and right inner solutions indeed agree with the left and right inner limit, Eqs. (3.146) and (3.150), respectively. Note that the solution for $\lambda_y(t)$ in leading order of ϵ vanishes in all three cases of inner and left and right outer equations.

4.1.6. Composite solutions and solution for control

The composite solutions are the sum of inner and outer solutions minus the overlaps,

$$x_{\text{comp}}(t) = x_O(t), \tag{4.94}$$
$$y_{\text{comp}}(t) = y_O(t) + Y_L((t-t_0)/\epsilon) - y_{\text{init}} + Y_R((t_1-t)/\epsilon) - y_{\text{end}}, \tag{4.95}$$
$$\lambda_{\text{comp}}(t) = \lambda_O(t). \tag{4.96}$$

See Section 3.5 and (Bender and Orszag, 2010) for further information about composite solutions. Here, $x_O(t)$, $y_O(t)$, and $\lambda_O(t)$ are the outer solutions Eqs. (4.33)-(4.35), while $Y_L(\tau_L)$ and $Y_R(\tau_R)$ are the left and right inner solutions given by Eqs. (4.65) and (4.82), respectively. The constant y_{end} is defined as $y_{\text{end}} = y_O(t_1)$ and the expression for y_{init} is given by Eq. (4.85). Equations (4.94)-(4.96) are the approximate solution to leading order in ϵ for optimal trajectory tracking. As a result of the singular perturbation expansion, the leading order solution depends on ϵ itself. The solution does not depend on the specific choice of the nonlinear term $R(x,y)$. The sole remains left by the nonlinearities $R(x,y)$ and $b(x,y)$ governing the system dynamics are the inner solutions Y_L and Y_R which depend on the coupling function $b(x,y)$.

The general expression for the control in terms of the state components $x(t)$ and $y(t)$ is

$$u(t) = \frac{1}{b(x(t), y(t))} \left(\dot{y}(t) - R(x(t), y(t)) \right). \tag{4.97}$$

In terms of the composite solutions x_{comp} and y_{comp}, $u(t)$ is

$$u(t) = \frac{1}{b(x_{\text{comp}}(t), y_{\text{comp}}(t))} \left(\dot{y}_{\text{comp}}(t) - R(x_{\text{comp}}(t), y_{\text{comp}}(t)) \right). \tag{4.98}$$

To obtain a result consistent with the approximate solution for state and co-state, Eq. (4.98) is expanded up to leading order in ϵ. The outer limit, valid for times $t_0 < t < t_1$, yields the identities

$$\lim_{\epsilon\to 0} Y_L\left(\left(t-t_0\right)/\epsilon\right) = y_{\text{init}}, \qquad \lim_{\epsilon\to 0} Y_R\left(\left(t_1-t\right)/\epsilon\right) = y_{\text{end}}, \tag{4.99}$$

$$\lim_{\epsilon\to 0} Y_L'\left(\left(t-t_0\right)/\epsilon\right) = 0, \qquad \lim_{\epsilon\to 0} Y_R'\left(\left(t_1-t\right)/\epsilon\right) = 0, \tag{4.100}$$

and the outer control signal as

$$u_O(t) = \lim_{\epsilon\to 0} u(t) = \frac{1}{b\left(x_O(t), y_O(t)\right)}\left(\dot{y}_O(t) - R\left(x_O(t), y_O(t)\right)\right). \tag{4.101}$$

The inner limits of the control are defined by

$$u\left(t_0+\epsilon\tau_L\right) = U_L\left(\tau_L\right) + \text{h.o.t.}, \tag{4.102}$$

$$u\left(t_1-\epsilon\tau_R\right) = U_R\left(\tau_R\right) + \text{h.o.t.}. \tag{4.103}$$

The abbreviation h.o.t. stands for higher order terms which vanish as $\epsilon \to 0$. The left and right outer control signals U_L and U_R depend on the rescaled times $\tau_L = \left(t-t_0\right)/\epsilon$ and $\tau_R = \left(t_1-t\right)/\epsilon$, respectively.

To compute U_L and U_R, the derivative $\dot{y}_{\text{comp}}$ must be expanded in ϵ,

$$\begin{aligned}\dot{y}_{\text{comp}}\left(t_0+\epsilon\tau_L\right) &= \dot{y}_O\left(t_0+\epsilon\tau_L\right) + \frac{1}{\epsilon}Y_L'\left(\tau_L\right) - \frac{1}{\epsilon}Y_R'\left(\left(t_1-t_0\right)/\epsilon - \tau_L\right)\\ &= \dot{y}_O\left(t_0\right) + \frac{1}{\epsilon}Y_L'\left(\tau_L\right) + \text{h.o.t.},\end{aligned} \tag{4.104}$$

$$\begin{aligned}\dot{y}_{\text{comp}}\left(t_1-\epsilon\tau_R\right) &= \dot{y}_O\left(t_1-\epsilon\tau_R\right) + \frac{1}{\epsilon}Y_L'\left(\left(t_1-t_0\right)/\epsilon - \tau_R\right) - \frac{1}{\epsilon}Y_R'\left(\tau_R\right)\\ &= \dot{y}_O\left(t_1\right) - \frac{1}{\epsilon}Y_R'\left(\tau_R\right) + \text{h.o.t.}.\end{aligned} \tag{4.105}$$

The expressions $\frac{1}{\epsilon}Y_R'\left(\left(t_1-t_0\right)/\epsilon - \tau_L\right)$ and $\frac{1}{\epsilon}Y_L'\left(\left(t_1-t_0\right)/\epsilon - \tau_R\right)$ are assumed to approach zero sufficiently fast as $\epsilon \to 0$. The terms proportional to $1/\epsilon$ diverge as $\epsilon \to 0$. This forbids a straightforward computation of $\lim_{\epsilon\to 0}$, and is the reason for the h.o.t. notation. The left and right inner limits of the control signal are obtained

as

$$\begin{aligned} U_L(\tau_L) &= \frac{1}{b(x_0, Y_L(\tau_L))}\left(\dot{y}_O(t_0) + \frac{1}{\epsilon}Y_L'(\tau_L) - R(x_0, Y_L(\tau_L))\right) \\ &= \frac{1}{b(x_0, Y_L(\tau_L))}\left(\dot{y}_O(t_0) - R(x_0, Y_L(\tau_L))\right) \\ &\quad + \frac{\sqrt{s_2}}{\epsilon}\operatorname{sign}(b(x_0, Y_L(\tau_L)))(y_{\text{init}} - Y_L(\tau_L)), \end{aligned} \tag{4.106}$$

$$\begin{aligned} U_R(\tau_R) &= \frac{1}{b(x_1, Y_R(\tau_R))}\left(\dot{y}_O(t_1) - \frac{1}{\epsilon}Y_R'(\tau_R) - R(x_1, Y_R(\tau_R))\right) \\ &= \frac{1}{b(x_1, Y_R(\tau_R))}\left(\dot{y}_O(t_1) - R(x_1, Y_R(\tau_R))\right) \\ &\quad - \frac{\sqrt{s_2}}{\epsilon}\operatorname{sign}(b(x_1, Y_R(\tau_R)))(y_{\text{end}} - Y_R(\tau_R)). \end{aligned} \tag{4.107}$$

Equation (4.65) is used to substitute $Y_L'(\tau_L)$, and analogously for $Y_R'(\tau_R)$.

The left and right matching conditions

$$u_O(t_0) = \lim_{\tau_L \to \infty} U_L(\tau_L), \qquad u_O(t_1) = \lim_{\tau_R \to \infty} U_R(\tau_R), \tag{4.108}$$

are satisfied because $Y_L'(\tau_L)$ and $Y_R'(\tau_R)$ approach zero as $\tau_L \to \infty$ and $\tau_R \to \infty$. The overlaps are obtained as

$$\begin{aligned} u_O(t_0) &= \frac{1}{b(x_O(t_0), y_O(t_0))}\left(\dot{y}_O(t_0) - R(x_O(t_0), y_O(t_0))\right) \\ &= \frac{1}{b(x_0, y_{\text{init}})}\left(\dot{y}_d(t_0) + \frac{a_2 s_1}{s_2}(x_0 - x_d(t_0)) + a_1(y_d(t_0) - y_{\text{init}}) - R(x_0, y_{\text{init}})\right), \end{aligned} \tag{4.109}$$

$$\begin{aligned} u_O(t_1) &= \frac{1}{b(x_O(t_1), y_O(t_1))}\left(\dot{y}_O(t_1) - R(x_O(t_1), y_O(t_1))\right) \\ &= \frac{1}{b(x_1, y_{\text{end}})}\left(\dot{y}_d(t_1) + \frac{a_2 s_1}{s_2}(x_1 - x_d(t_1)) + a_1(y_d(t_1) - y_{\text{end}}) - R(x_1, y_{\text{end}})\right). \end{aligned} \tag{4.110}$$

The differential equation for y_O, Eq. (4.28), is used to eliminate $\dot{y}(t_0)$ and $\dot{y}(t_1)$. Finally, the composite solution for the control signal is

$$u_{\text{comp}}(t) = u_O(t) + U_L((t - t_0)/\epsilon) + U_R((t_1 - t)/\epsilon) - u_O(t_0) - u_O(t_1). \tag{4.111}$$

The outer control signal $u_O(t)$ is given by Eq. (4.101), while the left and right inner control signals U_L and U_R are given by Eqs. (4.106) and (4.107), respectively. The long explicit expression for Eq. (4.111) is not written down explicitly.

4.1.7. The limit $\epsilon \to 0$

For $\epsilon = 0$, the analytical approximations derived in this section become exact. The exact solution displays a discontinuous state trajectory and a diverging control signal. The boundary layers of the state component $y(t)$ degenerates to a jump located at the beginning and the end of the time interval,

$$\begin{aligned}
\lim_{\epsilon\to 0} y(t) &= \lim_{\epsilon\to 0} y_{\text{comp}}(t) \\
&= y_O(t) + \lim_{\epsilon\to 0} Y_L\left((t-t_0)/\epsilon\right) - y_{\text{init}} + \lim_{\epsilon\to 0} Y_R\left((t_1-t)/\epsilon\right) - y_{\text{end}} \\
&= \begin{cases} y_O(t_0) + Y_L(0) - y_{\text{init}}, & t = t_0, \\ y_O(t), & t_0 < t < t_1, \\ y_O(t_1) + Y_R(0) - y_{\text{end}}, & t = t_1, \end{cases} \\
&= \begin{cases} y_0, & t = t_0, \\ y_O(t), & t_0 < t < t_1, \\ y_1, & t = t_1. \end{cases}
\end{aligned} \tag{4.112}$$

The state component $x(t)$ as well as the co-state $\lambda_x(t)$ do not exhibit boundary layers. Their solutions are continuous also for $\epsilon = 0$ and simply given by the outer solutions Eqs. (4.94) and (4.96). The remaining co-state $\lambda_y(t)$ vanishes identically, $\lambda_y(t) = 0$. Although $y_O(t)$ depends on the matching constants y_{init} and y_{end}, both constants are given solely in terms of the outer solutions and the initial and terminal conditions. Thus, to determine the height and the position of the jumps, it is not necessary to know any details about the dynamics of the boundary layers. For $\epsilon = 0$, no trace of the boundary layers is left in the composite solution except for the mere existence of the jumps. In particular, while the form of the boundary layers depends on the specific choice of the coupling function $b(x, y)$, the solution becomes independent of the coupling function $b(x, y)$ for $\epsilon = 0$. Thus, for $\epsilon = 0$, both possible sources of nonlinear system dynamics, the nonlinearity $R(x, y)$ and the coupling function $b(x, y)$, do entirely disappear from the solution for the controlled state trajectory.

To obtain the control signal for $\epsilon = 0$ from Eq. (4.111), the expression $Y_L'\left((t-t_0)/\epsilon\right)/\epsilon$ must be analyzed in the limit $\epsilon \to 0$. To that end, let the function $\delta_\epsilon(t)$ be defined as

$$\begin{aligned}
\delta_\epsilon(t) &= \begin{cases} \dfrac{1}{2\epsilon}\dfrac{1}{(y_{\text{init}} - y_0)} Y_L'\left(t/\epsilon\right), & t \geq 0, \\ \dfrac{1}{2\epsilon}\dfrac{1}{(y_{\text{init}} - y_0)} Y_L'\left(-t/\epsilon\right), & t < 0, \end{cases} \\
&= \frac{1}{2\epsilon}\frac{1}{(y_{\text{init}} - y_0)} Y_L'\left(|t|/\epsilon\right).
\end{aligned} \tag{4.113}$$

Note that $\delta_\epsilon(t)$ is continuous across $t = 0$ for all $\epsilon > 0$. A few computations prove

that $\delta_\epsilon(t)$ is a representation of the Dirac delta function as $\epsilon \to 0$,

$$\lim_{\epsilon\to 0} \delta_\epsilon(t) = \delta(t). \tag{4.114}$$

Indeed, from the differential equation for Y_L, Eq. (4.65) follows

$$\lim_{\epsilon\to 0} \frac{1}{\epsilon} Y_L'(0) = \lim_{\epsilon\to 0} \frac{1}{\epsilon} \sqrt{s_2}\,(y_{\text{init}} - Y_L(0))\,|b(x_0, Y_L(0))| = \text{sign}(y_{\text{init}} - y_0)\,\infty, \tag{4.115}$$

and therefore

$$\lim_{\epsilon\to 0} \delta_\epsilon(t) = \begin{cases} 0, & |t| > 0, \\ \infty, & t = 0. \end{cases} \tag{4.116}$$

It remains to show that

$$\int_{-\infty}^{\infty} dt \delta_\epsilon(t) = 1 \tag{4.117}$$

for all ϵ. Together with the substitutions $t_1 = -\epsilon\tau_L$ and $t_2 = \epsilon\tau_L$ and the initial and matching condition for Y_L, the integral over $\delta_\epsilon(t)$ yields

$$\begin{aligned} \int_{-\infty}^{\infty} dt \delta_\epsilon(t) &= \int_{-\infty}^{0} dt_1 \delta_\epsilon(t_1) + \int_{0}^{\infty} dt_2 \delta_\epsilon(t_2) \\ &= \frac{1}{2\epsilon} \frac{1}{(y_{\text{init}} - y_0)} \left(\int_{-\infty}^{0} dt_1 Y_L'(-t_1/\epsilon) + \int_{0}^{\infty} dt_2 Y_L'(t_2/\epsilon) \right) \\ &= \frac{1}{2} \frac{1}{(y_{\text{init}} - y_0)} \left(\int_{0}^{\infty} d\tau_L Y_L'(\tau_L) + \int_{0}^{\infty} d\tau_L Y_L'(\tau_L) \right) = 1. \end{aligned} \tag{4.118}$$

The result Eq. (4.118) is independent of the value of ϵ. Thus, Eq. (4.114) can be used to establish the identity

$$\lim_{\epsilon\to 0} \frac{1}{\epsilon} Y_L'((t - t_0)/\epsilon) = 2(y_{\text{init}} - y_0)\,\delta(t - t_0),\ t \geq t_0. \tag{4.119}$$

An analogous procedure applied to $Y_R'(\tau_R)$ yields the analogous result

$$\lim_{\epsilon\to 0} \frac{1}{\epsilon} Y_R'((t_1 - t)/\epsilon) = 2(y_{\text{end}} - y_1)\,\delta(t_1 - t),\ t_1 \geq t. \tag{4.120}$$

Equations (4.119) and (4.120) are used to compute the limits

$$\begin{aligned} &\lim_{\epsilon\to 0} U_L((t - t_0)/\epsilon) \\ =&\lim_{\epsilon\to 0} \frac{1}{b(x_0, Y_L((t - t_0)/\epsilon))} \left(\dot{y}_O(t_0) + \frac{1}{\epsilon} Y_L'((t - t_0)/\epsilon) - R(x_0, Y_L((t - t_0)/\epsilon)) \right) \\ =&\begin{cases} \frac{1}{b(x_0, y_{\text{init}})} (\dot{y}_O(t_0) - R(x_0, y_{\text{init}})) = u_O(t_0), & t_0 < t \leq t_1, \\ \frac{1}{b(x_0, y_0)} (\dot{y}_O(t_0) + 2(y_{\text{init}} - y_0)\,\delta(t - t_0) - R(x_0, y_0)), & t = t_0, \end{cases} \end{aligned} \tag{4.121}$$

and

$$\begin{aligned}
&\lim_{\epsilon\to 0} U_R\left(\left(t_1-t\right)/\epsilon\right)\\
=&\lim_{\epsilon\to 0}\frac{1}{b\left(x_1, Y_R\left(\left(t_1-t\right)/\epsilon\right)\right)}\left(\dot{y}_O\left(t_1\right)-\frac{1}{\epsilon}Y_R'\left(\left(t_1-t\right)/\epsilon\right)-R\left(x_1, Y_R\left(\left(t_1-t\right)/\epsilon\right)\right)\right)\\
=&\begin{cases}\frac{1}{b(x_1,y_{\text{end}})}\left(\dot{y}_O\left(t_1\right)-R\left(x_1,y_{\text{end}}\right)\right)=u_O\left(t_1\right), & t_0\le t<t_1,\\ \frac{1}{b(x_1,y_1)}\left(\dot{y}_O\left(t_1\right)-2\left(y_{\text{end}}-y_1\right)\delta\left(t_1-t\right)-R\left(x_1,y_1\right)\right), & t=t_1,\end{cases}
\end{aligned}\tag{4.122}$$

respectively. Finally, the exact solution for the control signal is

$$\begin{aligned}
\lim_{\epsilon\to 0} u\left(t\right) &= \lim_{\epsilon\to 0} u_{\text{comp}}\left(t\right)\\
&= u_O\left(t\right)+\lim_{\epsilon\to 0} U_L\left(\left(t-t_0\right)/\epsilon\right)-u_O\left(t_0\right)+\lim_{\epsilon\to 0} U_R\left(\left(t_1-t\right)/\epsilon\right)-u_O\left(t_1\right)\\
&= \begin{cases}\frac{1}{b(x_0,y_0)}\left(\dot{y}_O\left(t_0\right)+2\left(y_{\text{init}}-y_0\right)\delta\left(t-t_0\right)-R\left(x_0,y_0\right)\right), & t=t_0,\\ \frac{1}{b(x_O(t),y_O(t))}\left(\dot{y}_O\left(t\right)-R\left(x_O\left(t\right),y_O\left(t\right)\right)\right), & t_0<t<t_1,\\ \frac{1}{b(x_1,y_1)}\left(\dot{y}_O\left(t_1\right)-2\left(y_{\text{end}}-y_1\right)\delta\left(t_1-t\right)-R\left(x_1,y_1\right)\right), & t=t_1.\end{cases}
\end{aligned}\tag{4.123}$$

In contrast to the controlled state trajectory, the solution for the control signal depends on both possible nonlinearities, the coupling function $b\left(x,y\right)$ and the nonlinearity $R\left(x,y\right)$.

A discussion of the results can be found at the end of the next section.

4.2. Comparison with numerical results

4.2.1. Results

Numerical computations are performed with the ACADO Toolkit (Houska et al., 2011a,b), an open source program package for solving optimal control problems. Typically, a problem is solved on a time interval of length 1 with a time step width of $\Delta t = 10^{-3}$. The numerical computation of such an example takes about 20-30min with a standard Laptop. Computation time increases quickly with decreasing step width or increasing length of the time interval. For comparison with analytical solutions, the numerical result provided by ACADO is imported in Mathematica (Wolfram Research, Inc., 2014) and interpolated. Two example systems are investigated. Both are covered by the general analytical result from Section 4.1. The activator-controlled FHN model in is discussed in Example 4.2. Example 4.3 presents results for a mechanical system, namely the damped mathematical pendulum.

Example 4.2: Activator-controlled FHN model

See Example 1.2 for an introduction to the model and parameter values. The state equations are repeated here for convenience,

$$\begin{pmatrix} \dot{x}(t) \\ \dot{y}(t) \end{pmatrix} = \begin{pmatrix} a_0 + a_1 x(t) + a_2 y(t) \\ y(t) - \frac{1}{3} y(t)^3 - x(t) \end{pmatrix} + \begin{pmatrix} 0 \\ b(x(t), y(t)) \end{pmatrix} u(t). \tag{4.124}$$

In contrast to Example 1.4 with $b(x, y) = 1$, here a state-dependent coupling function b is assumed,

$$b(x, y) = \frac{11}{4} + x^2. \tag{4.125}$$

The small regularization parameter ϵ is set to

$$\epsilon = 10^{-3}, \tag{4.126}$$

which results in a regularization term with coefficient $\frac{1}{2}\epsilon^2 = 0.5 \times 10^{-6}$, see Eq. (4.3). The desired reference trajectory is an ellipse,

$$x_d(t) = A_x \cos(2\pi t/T) - \frac{1}{2}, \qquad y_d(t) = A_y \sin(2\pi t/T) + \frac{1}{2}, \tag{4.127}$$

with $A_x = 1$, $A_y = 15$, and $T = 1$. Within the time interval $0 = t_0 \leq t < t_1 = 1$, the controlled state trajectory shall follow the ellipse as closely as possible. The initial and terminal state lie on the desired trajectory,

$$x(t_0) = x_d(t_0) = \frac{1}{2}, \qquad y(t_0) = y_d(t_1) = \frac{15}{2}, \tag{4.128}$$

$$x(t_1) = x_d(t_1) = \frac{1}{2}, \qquad y(t_1) = y_d(t_1) = \frac{15}{2}. \tag{4.129}$$

Figure 4.1 compares the prescribed desired trajectory $\boldsymbol{x}_d(t)$ as given by Eq. (4.127) (blue solid line) with the numerically obtained optimally controlled state trajectory $\boldsymbol{x}_{num}(t)$ (red dashed line). While the controlled activator (Fig. 4.1 right) looks similar to the desired trajectory except for a constant shift, the controlled inhibitor (Fig. 4.1 right) is way off. Although the initial and terminal conditions comply with the desired trajectory, the solution for the activator component $y(t)$ exhibits some very steep transients at the beginning and end of the time interval. These transients can be interpreted as boundary layers described by the inner solutions. While this example differs from Example 1.4 in the coupling function $b(x, y)$, its controlled state trajectories are very similar. Indeed, the analytical solutions predicts an effect of $b(x, y)$ restricted to the boundary layer region of the activator component $y(t)$. Consequently,

the solutions for different coupling functions $b(x, y)$ are essentially identical inside the time domain.

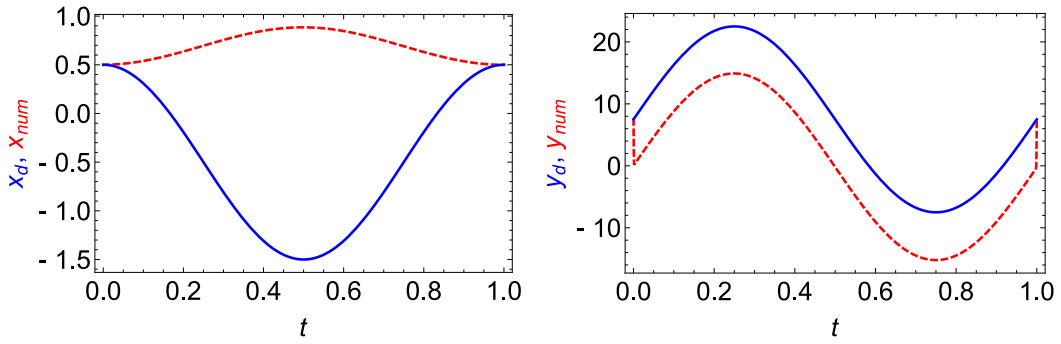

Figure 4.1.: Comparison of desired reference trajectory $\boldsymbol{x}_d(t)$ (blue solid line) and numerically obtained optimal trajectory $\boldsymbol{x}_{num}(t)$ (red dashed line) in the FHN model. The activator over time y (right) is similar to the reference trajectory in shape but shifted by an almost constant value, while the inhibitor over time x (left) is far off.

Figure 4.2 compares the analytical result for $\boldsymbol{x}(t)$ from Section 4.1 with its numerical counterpart. The agreement is excellent. Although only approximately valid, this demonstrates an astonishing accuracy of the analytical result for small values of ϵ. No difference between analytical and numerical is visible on this scale. Figure 4.3 visualizes $\boldsymbol{x}(t) - \boldsymbol{x}_{num}(t)$ and reveals relatively small differences in the bulk but somewhat larger differences close to the initial and terminal time, especially for the state component $y(t)$.

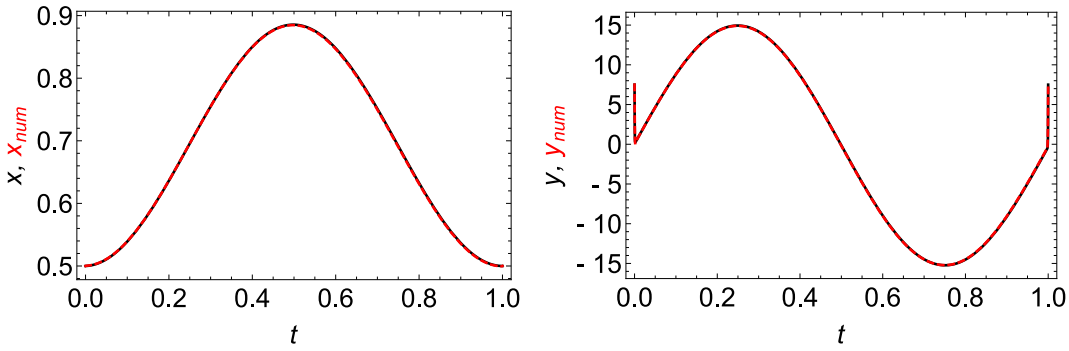

Figure 4.2.: Comparison of numerically obtained optimal trajectory (red dashed line) and analytical approximation (black solid line). On this scale, the agreement is perfect.

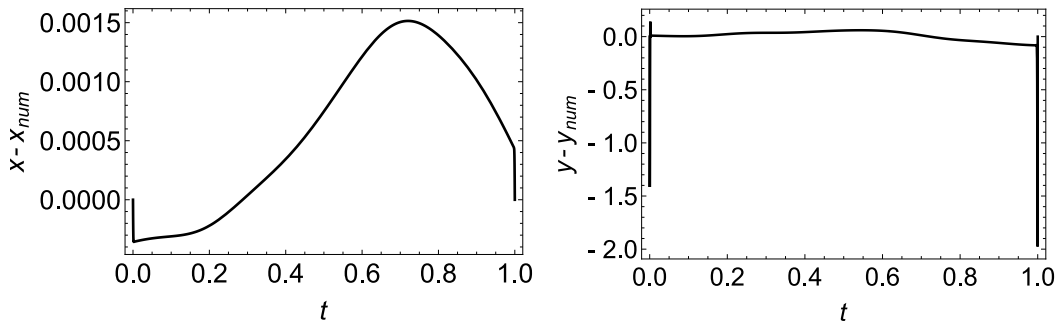

Figure 4.3.: Difference between analytical and numerical solution for inhibitor (left) and activator component (right) over time.

Zooming in on the initial (Fig. 4.4) and the terminal time (Fig. 4.5) uncovers the boundary layers. The activator (right) displays small deviations in the regions with the steepest slopes. This is certainly due to the limited temporal resolution of the numerical simulation. Note that the width of the boundary layers is approximately determined by the value of the regularization parameter ϵ. The value $\epsilon = 10^{-3}$ chosen for numerical simulations is identical to the temporal resolution of $\Delta t = 10^{-3}$. A result is the relatively large difference between analytical and numerical result at the initial and terminal times in Fig. 4.3. For values of ϵ in the range of the temporal resolution, the boundary layers cannot be resolved numerically with sufficient accuracy and result in discretization errors. Due to the computational cost of optimal control algorithms, decreasing the step width Δt is not really an option. Figure 4.4 and 4.5 left reveals a small boundary layer displayed only by the numerically obtained inhibitor component. This is not predicted by the analytical leading order approximation and results probably from higher order contributions of the perturbation expansion. Finally, note that the deviations between analytical and numerical result are slightly larger close to the terminal time (Fig. 4.5) than to the initial time (Fig. 4.4). This hints at the accumulation of numerical errors in the numerical result.

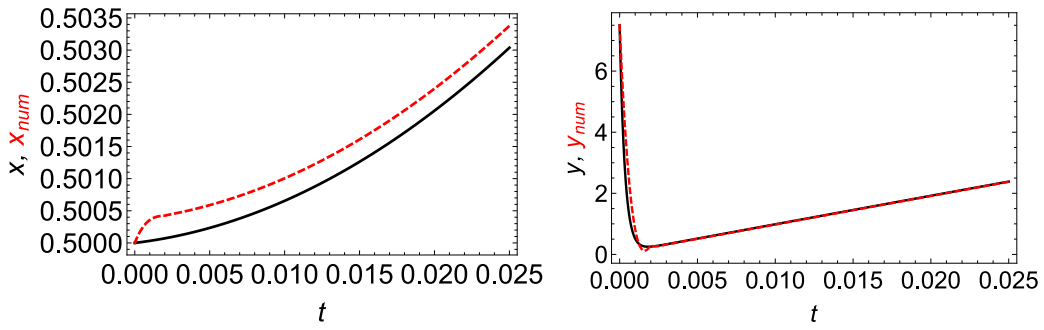

Figure 4.4.: Closeup of the left boundary layer for the activator component (right) shows perfect agreement between numerical (red dashed line) and analytical (black line) result except for the steepest slopes. The numerically obtained inhibitor component exhibits a very small boundary layer as well (left), while the leading order analytical result does not. Analytically, this boundary layer arises probably from contributions of higher order in ϵ.

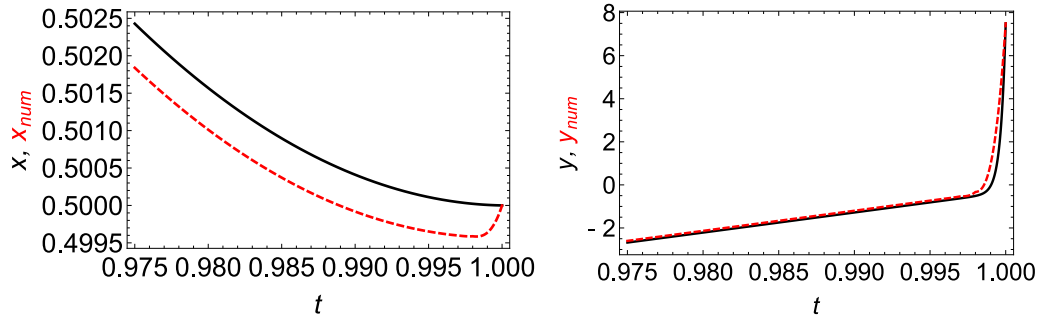

Figure 4.5.: Closeup of the right boundary layer for the activator. The agreement is worse than for the left boundary layer. This hints at the accumulation of numerical errors in the numerical result.

Figure 4.6 concludes with a comparison between analytical and numerical solution for the control. The control attains its largest values at the initial and terminal time. Analytically, these spikes approach the form of a Dirac delta distribution for a decreasing values of ϵ.

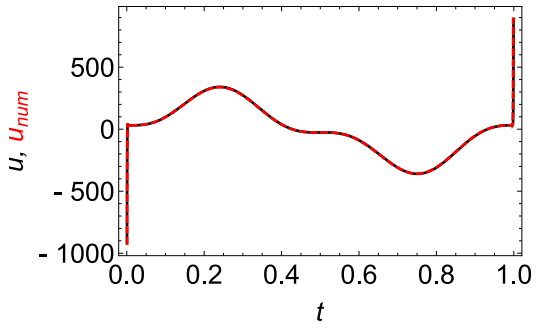

Figure 4.6.: Comparison of the numerically obtained control u (red dashed line) and its analytical approximation (black solid line). The numerical and analytical control solutions attain large values at the boundaries of the time domain due to the appearance of boundary layers in the state component $y\,(t)$.

Example 4.3: Mathematical Pendulum

The mathematical pendulum is an example for a mechanical control system, see Example 1.1. The controlled state equation is

$$\dot{x}\,(t) = y\,(t)\,, \tag{4.130}$$
$$\dot{y}\,(t) = -\gamma y\,(t) - \sin\,(x\,(t)) + u\,(t)\,. \tag{4.131}$$

For mechanical systems, the general analytical result from Section 4.1 simplifies considerably due to the fixed parameter values

$$a_0 = 0, \qquad a_1 = 0, \qquad a_2 = 1. \tag{4.132}$$

The desired trajectory $\boldsymbol{x}_d\,(t)$ is

$$x_d\,(t) = \cos\,(2\pi t)\,, \qquad y_d\,(t) = \cos\,(2\pi t) + \sin\,(4\pi t)\,. \tag{4.133}$$

The initial and terminal conditions

$$x\,(t_0) = -1, \qquad y\,(t_0) = -1, \qquad x\,(t_1) = -1, \qquad y\,(t_1) = -1, \tag{4.134}$$

do not comply with the desired trajectory. As before, the small parameter ϵ is

$$\epsilon = 10^{-3}. \tag{4.135}$$

Figure 4.7 compares the desired trajectory $\boldsymbol{x}_d\,(t)$ with the numerically obtained optimally controlled state trajectory $\boldsymbol{x}_{num}\,(t)$. The velocity $y\,(t)$ is much closer

to its desired counterpart than the position over time $x(t)$. Initial and terminal boundary layers occur for the velocity $y(t)$. Figure 4.8 compares the corresponding analytical result with the numerical solution and reveals almost perfect agreement.

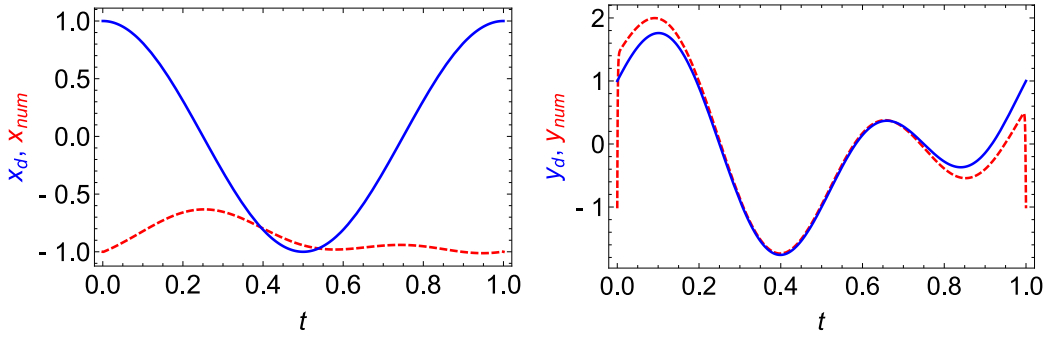

Figure 4.7.: Desired (blue solid line) and optimally controlled position x (left) and velocity y (right) over time for the damped pendulum. While the controlled velocity follows the desired velocity closely, the position is far off. The velocity exhibits an initial and terminal boundary layer.

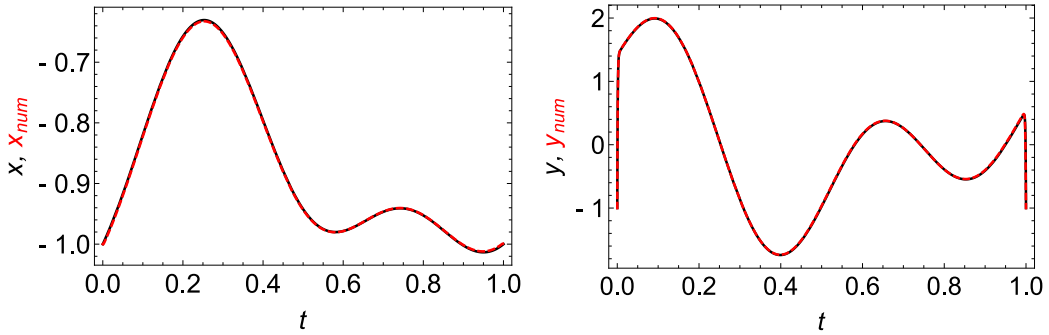

Figure 4.8.: Comparison of analytical (black solid line) and numerical (red dashed line) result for the controlled position (left) and velocity (right) over time reveals nearly perfect agreement.

As a second example, consider the desired trajectory

$$x_d(t) = \cos(2\pi t), \qquad y_d(t) = \cos(2\pi t) + \sin(4\pi t) + 100, \tag{4.136}$$

together with the same initial and terminal conditions Eqs. (4.134) as above. Equations (4.136) differ from the desired trajectory from Eqs. (4.133) only in a constant shift of the activator component, $y_d(t) \to y_d(t) + \alpha$. Figure 4.9 compares the numerical solution for the controlled state trajectory $\boldsymbol{x}_{num}(t)$ with the desired trajectory $\boldsymbol{x}_d(t)$ as given by Eq. (4.136). In contrast to Fig. 4.7, the controlled position over time (Fig. 4.9 left) is much closer to its desired counterpart than the velocity over time (Fig. 4.9 right). Figure 4.10 shows a direct comparison of the controlled state trajectories for the desired trajectory

Eqs. (4.133) (black solid line) and for the desired trajectory Eqs. (4.136) (yellow dotted line). Surprisingly, the controlled state trajectories are identical.

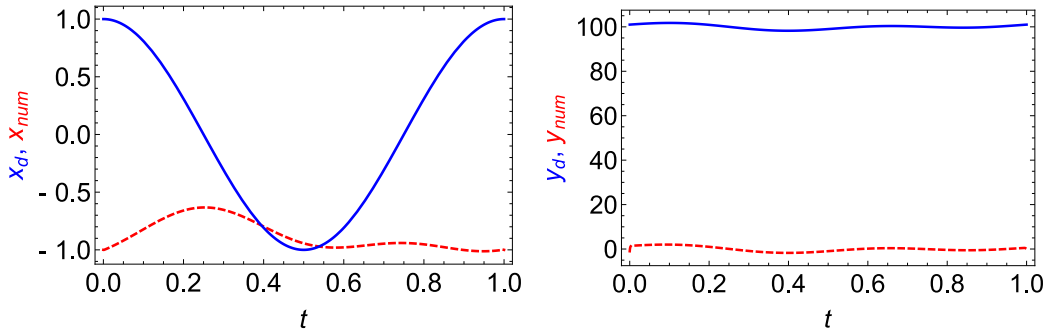

Figure 4.9.: Desired (blue solid line) and actually realized position x (left) and velocity y (right) over time for the damped pendulum for a desired trajectory given by Eqs. (4.136). The velocity over time differs significantly from its desired counterpart (right), while the position over time lies in the correct range of values (left).

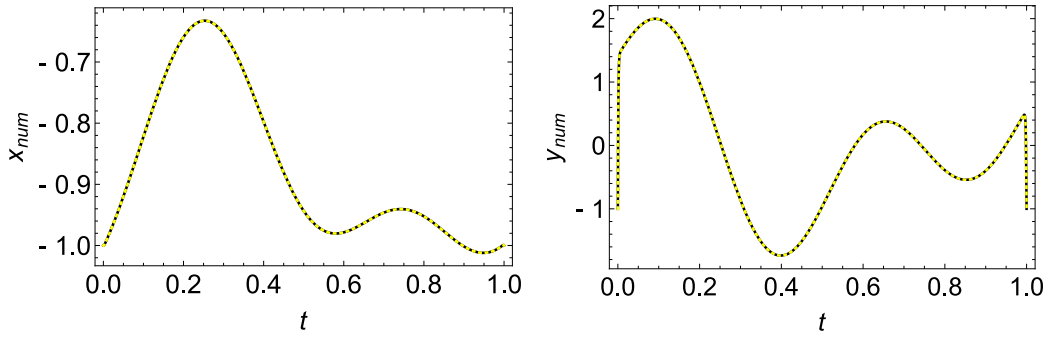

Figure 4.10.: Invariance of the optimal state trajectory under a constant shift of the desired velocity over time. Position (left) and velocity (right) over time of the controlled state trajectory for two desired trajectories, Eqs. (4.133) (black solid line) and Eqs. (4.136) (yellow dotted line). The desired trajectories differ by a constant shift of the desired velocity. Surprisingly, both controlled state trajectories are identical.

A more detailed look at the analytical result confirms the findings of Fig. 4.10. For mechanical systems, the outer solution is given by

$$\begin{pmatrix} x_O(t) \\ y_O(t) \end{pmatrix} = \mathbf{\Phi}(t, t_0) \begin{pmatrix} x_{\text{init}} \\ y_{\text{init}} \end{pmatrix} + \int_{t_0}^{t} d\tau \mathbf{\Phi}(t, \tau) \boldsymbol{f}(\tau), \tag{4.137}$$

with state transition matrix $\mathbf{\Phi}(t, t_0)$ and inhomogeneity

$$\boldsymbol{f}(t) = \begin{pmatrix} 0 \\ \dot{y}_d(t) - \frac{s_1}{s_2} x_d(t) \end{pmatrix}. \tag{4.138}$$

While $\boldsymbol{f}(t)$ depends on $x_d(t)$ and the time derivative $\dot{y}_d(t)$, it is independent of $y_d(t)$ itself. Thus, the inhomogeneity is invariant under a constant shift $y_d(t) \to y_d(t) + \alpha$. A more careful analysis reveals that the full composite solution for $\boldsymbol{x}(t)$, Eqs. (4.94) and (4.95), is independent of a constant shift α as long as $a_1 = 0$. Consequently, the composite solution for the control signal, Eq. (4.111), is independent of a constant shift α as well.

4.2.2. Discussion

Analytical approximations for optimal trajectory tracking for a two-dimensional dynamical system were derived in Section 4.1. The system includes the mechanical control systems from Example 1.1 and the activator-controlled FHN model from Example 1.2 as special cases. The control acts on the nonlinear equation (4.2) for $y(t)$ while the uncontrolled equation (4.1) for $x(t)$ is linear.

The necessary optimality conditions are rearranged such that the highest order time derivative as well as every occurrence of the nonlinearity $R(x, y)$ is multiplied by the small parameter ϵ. This constitutes a system of singularly perturbed differential equations amenable to a perturbative treatment. The solution reveals that the y-component, i.e., the activator of the FHN model or the velocity of mechanical systems, exhibits steep transition regions close to the initial and terminal time. In the context of singular perturbation theory, these transitions are interpreted as boundary layers with width ϵ and arise as solutions to the inner equations. The inner solutions connect the initial and terminal condition, respectively, with the outer solution. The outer solution is valid only within the time domain. Boundary layers occur even if the initial and terminal conditions lie on the desired trajectory. The outer equations are linear and their analytical solutions are available in closed form. The inner equations depend on the nonlinear coupling function $b(x, y)$. Neither the outer nor the inner equations depend on the nonlinearity $R(x, y)$. The nonlinearity $R(x, y)$ is entirely absorbed by the control signal. The control signal depends on $R(x, y)$ as well as on $b(x, y)$.

As $\epsilon \to 0$, the boundary layers degenerate to jumps located at the initial and terminal time. Simultaneously, the control signal diverges and approaches the form of a Dirac delta function. The strength of the delta kicks, i.e., the coefficient of the Dirac delta function, is twice the height of the jumps. Because the delta kick is located right at the time domain boundaries, only half of the kick contributes to the time evolution. For $\epsilon = 0$, the composite solution is an exact solution to optimal trajectory tracking. The analytical form of the composite solution is entirely determined by the outer equations. No traces of the boundary layers remain except for the mere existence of the jumps in $y(t)$. However, the existence of these jumps, and therefore the existence of the entire exact solution, relies on the existence of solutions to the inner equations for the appropriate initial, terminal, and matching conditions. In contrast to the perturbative result for $\epsilon > 0$, the dynamics is independent of the coupling

function $b(x, y)$. Both sources $R(x, y)$ and $b(x, y)$ of nonlinearity are irrelevant for the controlled state trajectory. It is in this sense that we are able to speak about linearity in unregularized nonlinear optimal control. This result unveils a linear structure underlying nonlinear optimal trajectory tracking.

A first analysis of the analytical results reveals that the controlled state $y(t)$ is invariant under a constant shift α of the desired velocity, $y_d(t) \to y_d(t) + \alpha$, as long as $a_1 = 0$. Note that $a_1 = 0$ for all mechanical control systems, see Eq. (4.1). This behavior is partially retained for the FHN model as long as a_1 is small. Such insights can hardly be obtained from numerical simulations alone. The impact of this finding depends on the physical interpretation of the dynamical system. For mechanical systems, $\dot{x}(t) = y(t)$ denotes the velocity of the system. Whereas shifting $y_d(t)$ has no effect on the controlled velocity $y(t)$, transforming the desired position over time as $x_d(t) \to x_d(t) + \alpha t$ also changes $y(t)$. The controlled velocity can nevertheless be affected by appropriately designing desired trajectories.

4.3. Optimal feedback control

The approximate solution to the necessary optimality conditions depends on the initial state $\boldsymbol{x}_0 = \boldsymbol{x}(t_0)$. Two possibilities exist to determine $\boldsymbol{x}_0$. Either the system is prepared in state $\boldsymbol{x}_0$, or $\boldsymbol{x}_0$ is obtained by measurement. Knowing the value of $\boldsymbol{x}_0$, no further information about the controlled system's state is necessary to compute the control signal and the state's time evolution. The control is an *open loop control*.

External influences not modeled by the system dynamics can destabilize the controlled system. A measurement $\tilde{x}_0 = x\left(\tilde{t}_0\right)$, $\tilde{y}_0 = y\left(\tilde{t}_0\right)$ performed at a later time $t = \tilde{t}_0 > t_0$ can be used to update the control with $x_0 = \tilde{x}_0$, $y_0 = \tilde{y}_0$ as the new initial condition. By feeding repeated measurements back into the controlled system, it is possible to counteract unmodeled perturbations occurring in between measurements. The optimal control solution derived in Section 4.1 is a *sampled-data feedback law* (Bryson and Ho, 1975). The initial time t_0 is the most recent sampling time, and the initial conditions $x_0 = x(t_0)$, $y_0 = y(t_0)$ are measurements of the controlled system's state.

If a continuous monitoring of the system's state is possible, one can set $t_0 \to t$ in the composite state solutions Eqs. (4.94)-(4.98). The initial conditions $x_0 = x(t)$, $y_0 = y(t)$ become functions of the current state of the controlled system itself. This is known as an optimal *continuous time feedback law,* also called a *closed loop control*. Similarly, setting $t_0 \to t - T$ and $x_0 = x(t - T)$, $y_0 = y(t - T)$ yields a *continuous time-delayed feedback law.* The state measurements are fed back to the system after a delay time $T > 0$.

Thus, optimal feedback control requires knowledge of the controlled state trajectory's dependence on its initial state $\boldsymbol{x}_0$. A numerical solution to optimal control, determined for a single specified value of $\boldsymbol{x}_0$, cannot be used for feedback control.

Instead, optimal feedback control is obtained from the Hamilton-Jacobi-Bellman equation. This PDE is the central object of the Dynamic Programming approach to optimal control founded by Richard Bellmann and coworkers (Bellman, 2003).

Consider the target functional

$$\mathcal{J}\left[\boldsymbol{x}\left(t\right),\boldsymbol{u}\left(t\right)\right]=\int_{t_0}^{t_1}dt\frac{1}{2}\left(\boldsymbol{x}\left(t\right)-\boldsymbol{x}_d\left(t\right)\right)^T\boldsymbol{\mathcal{S}}\left(\boldsymbol{x}\left(t\right)-\boldsymbol{x}_d\left(t\right)\right)$$
$$+\frac{1}{2}\left(\boldsymbol{x}\left(t_1\right)-\boldsymbol{x}_1\right)^T\boldsymbol{\mathcal{S}}_1\left(\boldsymbol{x}\left(t_1\right)-\boldsymbol{x}_1\right)+\frac{\epsilon^2}{2}\int_{t_0}^{t_1}dt\left(\boldsymbol{u}\left(t\right)\right)^2. \tag{4.139}$$

The functional $\mathcal{J}\left[\boldsymbol{x}\left(t\right),\boldsymbol{u}\left(t\right)\right]$ is to be minimized subject to the constraints

$$\dot{\boldsymbol{x}}\left(t\right)=\boldsymbol{R}\left(\boldsymbol{x}\left(t\right)\right)+\boldsymbol{\mathcal{B}}\left(\boldsymbol{x}\left(t\right)\right)\boldsymbol{u}\left(t\right),\qquad\qquad\boldsymbol{x}\left(t_0\right)=\boldsymbol{x}_0. \tag{4.140}$$

Denote the minimal value of $\mathcal{J}$ by $\mathcal{J}_0$. $\mathcal{J}_0$ is obtained by evaluating $\mathcal{J}$ at the optimally controlled state trajectory $\boldsymbol{x}\left(t\right)$ and its corresponding control signal $\boldsymbol{u}\left(t\right)$. $\mathcal{J}_0$ can be considered as a function of the initial state $\boldsymbol{x}_0$ and the initial time t_0, $\mathcal{J}_0=\mathcal{J}_0\left(\boldsymbol{x}_0,t_0\right)$. The Hamilton-Jacobi-Bellman equation is a nonlinear evolution equation for $\mathcal{J}_0$ given by (Bryson and Ho, 1975)

$$0=\min_{\boldsymbol{u}}\left\{\nabla\mathcal{J}_0\left(\boldsymbol{x},t\right)\left(\boldsymbol{R}\left(\boldsymbol{x}\right)+\boldsymbol{\mathcal{B}}\left(\boldsymbol{x}\right)\boldsymbol{u}\right)+\frac{1}{2}\left(\boldsymbol{x}-\boldsymbol{x}_d\left(t\right)\right)^T\boldsymbol{\mathcal{S}}\left(\boldsymbol{x}-\boldsymbol{x}_d\left(t\right)\right)+\frac{\epsilon^2}{2}\boldsymbol{u}^2\right\}$$
$$+\frac{\partial}{\partial t}\mathcal{J}_0\left(\boldsymbol{x},t\right). \tag{4.141}$$

Equation (4.141) is supplemented with the terminal condition

$$\mathcal{J}_0\left(\boldsymbol{x},t_1\right)=\frac{1}{2}\left(\boldsymbol{x}-\boldsymbol{x}_1\right)^T\boldsymbol{\mathcal{S}}_1\left(\boldsymbol{x}-\boldsymbol{x}_1\right). \tag{4.142}$$

The co-state $\boldsymbol{\lambda}\left(t\right)$, considered as a function of the initial state $\boldsymbol{x}_0$, is given by the gradient of $\mathcal{J}_0$,

$$\boldsymbol{\lambda}^T\left(t\right)=\nabla\mathcal{J}_0\left(\boldsymbol{x}_0,t\right). \tag{4.143}$$

Determining the minimum on the right hand side of Eq. (4.141) yields a relation between the control and the gradient of $\mathcal{J}_0$. This relation is analogous to the stationarity condition, Eq. (3.10), of the necessary optimality conditions,

$$\nabla\mathcal{J}_0\left(\boldsymbol{x},t\right)\boldsymbol{\mathcal{B}}\left(\boldsymbol{x}\right)+\epsilon^2\boldsymbol{u}^T=\boldsymbol{0}. \tag{4.144}$$

Solving for the control signal $\boldsymbol{u}$ yields

$$\boldsymbol{u}=-\frac{1}{\epsilon^2}\boldsymbol{\mathcal{B}}^T\left(\boldsymbol{x}\right)\nabla\mathcal{J}_0^T\left(\boldsymbol{x},t\right). \tag{4.145}$$

The Hamilton-Jacobi-Bellman equation for optimal trajectory tracking becomes

$$-\epsilon^2 \frac{\partial}{\partial t} \mathcal{J}_0 (\boldsymbol{x}, t) = -\frac{1}{2} \nabla \mathcal{J}_0 (\boldsymbol{x}, t) \boldsymbol{\mathcal{B}} (\boldsymbol{x}) \boldsymbol{\mathcal{B}}^T (\boldsymbol{x}) \nabla \mathcal{J}_0^T (\boldsymbol{x}, t) + \epsilon^2 \nabla \mathcal{J}_0 (\boldsymbol{x}, t) \boldsymbol{R} (\boldsymbol{x}) + \frac{\epsilon^2}{2} (\boldsymbol{x} - \boldsymbol{x}_d (t))^T \boldsymbol{\mathcal{S}} (\boldsymbol{x} - \boldsymbol{x}_d (t)) . \tag{4.146}$$

Knowing the solution to Eq. (4.146), the open loop control signal for a system with initial state $\boldsymbol{x} (t_0) = \boldsymbol{x}_0$ is recovered from Eq. (4.145) as

$$\boldsymbol{u} (t) = -\frac{1}{\epsilon^2} \boldsymbol{\mathcal{B}}^T (\boldsymbol{x}_0) \nabla \mathcal{J}_0^T (\boldsymbol{x}_0, t) . \tag{4.147}$$

A continuous time feedback law $\boldsymbol{u} (t) = \boldsymbol{u} (\boldsymbol{x} (t) , t)$ depends on the actual state $\boldsymbol{x} (t)$ of the controlled system and is given by

$$\boldsymbol{u} (t) = \boldsymbol{u} (\boldsymbol{x} (t) , t) = -\frac{1}{\epsilon^2} \boldsymbol{\mathcal{B}}^T (\boldsymbol{x} (t)) \nabla \mathcal{J}_0^T (\boldsymbol{x} (t) , t) . \tag{4.148}$$

We do not attempt to solve the nonlinear PDE Eq. (4.146), but end with some concluding remarks about the difficulties encountered in doing so. First of all, for a vanishing regularization parameter $\epsilon = 0$, Eq. (4.146) suffers from a similar degeneracy as the necessary optimality conditions. Because the time derivative $\partial_t \mathcal{J}_0$ vanishes for $\epsilon = 0$, $\mathcal{J}_0$ cannot satisfy the terminal condition (4.142). Second, to solve Eq. (4.146) numerically for $\epsilon > 0$ is a formidable task, especially if the dimension n of the state space is large. A discretization with N_x points for a single state space dimension results in N_x^n discretization points for the full state space. The computational cost increases exponentially with the state space dimension. This is the "curse of dimensionality", as it was called by Bellman himself (Bellman, 2003).

4.3.1. Continuous time feedback

4.3.1.1. Derivation of the feedback law

The approximate solution to the optimal trajectory tracking problem, Eqs. (4.94)-(4.96) and Eq. (4.111), is rendered as a continuous time feedback law. First, the initial conditions x_0 and y_0 are given by the controlled state components $x_0 = x (t_0) = x (t)$ and $y_0 = y (t_0) = y (t)$, respectively. Second, every explicit appearance of t_0 in Eqs. (4.94)-(4.96) and Eq. (4.111) is substituted by $t_0 \to t$. All constants which depend on time t_0, as e.g. x_{init}, become time dependent on the current time t. To minimize the confusion, these constants are written as

$$x_{\text{init}} = x_{\text{init}} (t) , \qquad y_{\text{init}} = y_{\text{init}} (t) , \qquad y_{\text{end}} = y_{\text{end}} (t) . \tag{4.149}$$

The outer solutions $x_O (t)$ and $y_O (t)$ given by Eqs. (4.33) and (4.34) assume a particularly simple form,

$$x_O (t) = x_{\text{init}} (t) , \qquad y_O (t) = y_{\text{init}} (t) , \tag{4.150}$$

and all integral terms vanish. The composite solutions Eqs. (4.94) and (4.95) reduce to

$$x_{\text{comp}}(t) = x_O(t) = x_{\text{init}}(t), \qquad y_{\text{comp}}(t) = Y_L(0) + Y_R((t_1 - t)/\epsilon) - y_{\text{end}}(t). \tag{4.151}$$

The composite control signal Eq. (4.111) becomes

$$u_{\text{comp}}(t) = U_L(0) + U_R((t_1 - t)/\epsilon) - u_O(t_1). \tag{4.152}$$

Note that Y_R, U_L, U_R as well as $u_O(t_1)$ are just abbreviations and still depend on time t through the time-dependent parameters from Eq. (4.149). The terms originating from the right boundary layer become important only for times $t \lesssim t_1$ close to the terminal time. For simplicity, consider the limit $t_1 \to \infty$. Because of

$$\lim_{t_1 \to \infty} Y_R((t_1 - t)/\epsilon) = y_{\text{end}}, \tag{4.153}$$

$$\lim_{t_1 \to \infty} U_R((t_1 - t)/\epsilon) = \lim_{t_1 \to \infty} u_O(t_1), \tag{4.154}$$

the terms originating from the right boundary layer in Eqs. (4.151), (4.152) cancel. Together with $x_{\text{init}}(t) = x_0 = x(t)$ and $Y_L(0) = y_0 = y(t)$, the composite state is

$$x_{\text{comp}}(t) = x_0 = x(t), \qquad y_{\text{comp}}(t) = y_0 = y(t). \tag{4.155}$$

The composite control solution simplifies to

$$u_{\text{comp}}(t) = U_L(0). \tag{4.156}$$

Note that $U_L(0)$ still depends on time t through the constants $x_0 = x(t)$, $y_0 = y(t)$ and $y_{\text{init}}^{\infty}(t)$. Here, $y_{\text{init}}^{\infty}(t)$ denotes the constant $y_{\text{init}}(t)$ in the limit $t_1 \to \infty$ given by

$$\begin{aligned} y_{\text{init}}^{\infty}(t) = \lim_{t_1 \to \infty} y_{\text{init}}(t) = \int_t^{\infty} e^{(t-\tau)\varphi_1} \left(\frac{a_2 s_1}{s_2} x_d(\tau) - (a_1 + \varphi_1) y_d(\tau) \right) d\tau \\ - \frac{1}{a_2}(\varphi_1 + a_1) x(t) - \frac{a_0}{a_2}\left(\frac{a_1}{\varphi_1} + 1 \right) + y_d(t). \end{aligned} \tag{4.157}$$

All occurrences of x_0 and y_0 are substituted with $x(t)$ and $y(t)$, respectively. The feedback control derived by the outlined procedure is called $u_{\text{fb}}(t)$ and given by

$$\begin{aligned} u_{\text{fb}}(t) = U_L(0) &= \frac{1}{b(x_0, y_0)} \left(\dot{y}_O(t_0) + \frac{1}{\epsilon} Y_L'(0) - R(x_0, y_0) \right) \\ &= \frac{1}{b(x(t), y(t))} \left(a_1 (y_d(t) - y_{\text{init}}^{\infty}(t)) + \dot{y}_d(t) + \frac{a_2 s_1}{s_2} (x(t) - x_d(t)) \right) \\ &+ \frac{1}{b(x(t), y(t))} \left(\frac{1}{\epsilon} \sqrt{s_2} (y_{\text{init}}^{\infty}(t) - y(t)) |b(x(t), y(t))| - R(x(t), y(t)) \right). \end{aligned} \tag{4.158}$$

Equations (4.65) and (4.28) were used to substitute the expression $Y_L'(0)$ and $\dot{y}_O(t)$, respectively. The feedback law Eq. (4.158) explicitly depends on the nonlinearities $R(x, y)$ and $b(x, y)$ and is only valid for an infinite terminal time $t_1 \to \infty$. An analogous but more complicated and longer expression can be derived for finite time intervals $t_0 \leq t \leq t_1$ from Eq. (4.152).

4.3.1.2. Feedback-controlled state trajectory

The time evolution of the controlled state trajectory under feedback control Eq. (4.158) is analyzed. Using Eq. (4.158) in the controlled state equations (4.1) and (4.2) results in

$$\dot{x}(t) = y(t), \tag{4.159}$$

$$\begin{aligned}\dot{y}(t) &= R(x(t), y(t)) + b(x(t), y(t))\, u_{\mathrm{fb}}(t) \\ &= a_1 (y_d(t) - y_{\mathrm{init}}^{\infty}(t)) + \dot{y}_d(t) + \frac{a_2 s_1}{s_2} (x(t) - x_d(t)) \\ &+ \frac{1}{\epsilon} \sqrt{s_2}\, (y_{\mathrm{init}}^{\infty}(t) - y(t))\, |b(x(t), y(t))|.\end{aligned} \tag{4.160}$$

Note that the nonlinearity $R(x, y)$ is eliminated from Eq. (4.160), whereas the dependence on the coupling function $b(x, y)$ is retained. Note that $y_{\mathrm{init}}^{\infty}(t)$ as given by Eq. (4.157) depends on $x(t)$ as well. Equations (4.159) and (4.160) have to be solved with the initial conditions

$$x(t_0) = x_0^{\mathrm{fb}}, \qquad y(t_0) = y_0^{\mathrm{fb}}. \tag{4.161}$$

Due to the coupling function $b(x, y)$, Eqs. (4.159) and (4.160) are nonlinear. No exact analytical closed form solution exists. However, Eqs. (4.159) and (4.160) can be solved perturbatively using the small parameter ϵ for a perturbation expansion. Due to the appearance of $1/\epsilon$, Eqs. (4.159) and (4.160) constitute a singularly perturbed system of differential equations.

A procedure analogous to the approach to open loop control is applied. The inner and outer equations and their solutions must be determined. Combining them to a composite solution yields an approximate solution uniformly valid over the while time interval. An initial boundary layer is expected close to the initial time t_0. Because of the assumed infinite terminal time $t_1 \to \infty$, no terminal boundary layer exists. The outer variables are denoted with index O,

$$x(t) = x_O(t), \qquad y(t) = y_O(t). \tag{4.162}$$

To leading order in ϵ, the outer equations are

$$\dot{x}_O(t) = y_O(t), \tag{4.163}$$

$$0 = (y_{\mathrm{init}}^{\infty}(t) - y_O(t))\, |b(x_O(t), y_O(t))|. \tag{4.164}$$

Because b does not have a root by assumption, the unique solution for $y_O(t)$ is

$$y_O(t) = y_{\mathrm{init}}^{\infty}(t). \tag{4.165}$$

The expression for $y_{\text{init}}^{\infty}(t)$ still depends on $x(t)$, see Eq. (4.157). The state $x_O(t)$ is governed by the differential equation

$$\dot{x}_O(t) = -\frac{1}{a_2}(\varphi_1 + a_1) x_O(t) - \frac{a_0}{a_2}\left(\frac{a_1}{\varphi_1} + 1\right) + y_d(t) + \int_t^{\infty} e^{(t-\tau)\varphi_1}\left(\frac{a_2 s_1}{s_2} x_d(\tau) - (a_1 + \varphi_1) y_d(\tau)\right) d\tau. \tag{4.166}$$

Equation (4.166) must be solved with the initial condition

$$x(t_0) = x_{\text{init}}^{\text{fb}}. \tag{4.167}$$

The constant $x_{\text{init}}^{\text{fb}}$ has to be determined by matching the outer solutions with the inner solutions. The solution for $x_O(t)$ is given by

$$x_O(t) = x_{\text{init}}^{\text{fb}} \exp\left(-(t - t_0)\frac{(a_1 + \varphi_1)}{a_2}\right) + \int_{t_0}^{t} g\left(\tilde{t}\right) \exp\left(\frac{(a_1 + \varphi_1)}{a_2}\left(\tilde{t} - t\right)\right) d\tilde{t}, \tag{4.168}$$

with abbreviation $g(t)$

$$g(t) = -\frac{a_0}{a_2}\left(\frac{a_1}{\varphi_1} + 1\right) + y_d(t) + \int_t^{\infty} e^{(t-\tau)\varphi_1}\left(\frac{a_2 s_1}{s_2} x_d(\tau) - (a_1 + \varphi_1) y_d(\tau)\right) d\tau. \tag{4.169}$$

The outer equation are not able to satisfy both initial conditions Eqs. (4.161). The initial boundary layer is resolved using the time scale

$$\tau_L = (t - t_0)/\epsilon \tag{4.170}$$

and rescaled inner solutions

$$X_L(\tau_L) = X_L((t - t_0)/\epsilon) = x(t) = x(t_0 + \epsilon\tau_L), \tag{4.171}$$

$$Y_L(\tau_L) = Y_L((t - t_0)/\epsilon) = y(t) = y(t_0 + \epsilon\tau_L). \tag{4.172}$$

Rewritten with the new time scale and rescaled functions, the feedback-controlled state equations (4.159), (4.160) are

$$\frac{1}{\epsilon}\dot{X}_L(\tau_L) = Y(\tau_L), \tag{4.173}$$

$$\frac{1}{\epsilon}\dot{Y}_L(\tau_L) = a_1\left(y_d(t_0 + \epsilon\tau_L) - y_{\text{init}}^{\infty}(t_0 + \epsilon\tau_L)\right) + \dot{y}_d(t_0 + \epsilon\tau_L) + \frac{a_2 s_1}{s_2}\left(X_L(\tau_L) - x_d(t_0 + \epsilon\tau_L)\right) + \frac{1}{\epsilon}\sqrt{s_2}\left(y_{\text{init}}^{\infty}(t_0 + \epsilon\tau_L) - Y_L(\tau_L)\right)\left|b\left(x_L(\tau_L), Y_L(\tau_L)\right)\right|. \tag{4.174}$$

Note that $y_{\text{init}}^{\infty}(t_0+\epsilon\tau_L)$ still depends on $x(t)$ and becomes

$$y_{\text{init}}^{\infty}(t_0+\epsilon\tau_L) = \int_{t_0+\epsilon\tau_L}^{\infty} e^{(t_0+\epsilon\tau_L-\tau)\varphi_1}\left(\frac{a_2 s_1}{s_2}x_d(\tau)-(a_1+\varphi_1)\,y_d(\tau)\right)d\tau \\ -\frac{1}{a_2}(\varphi_1+a_1)\,X_L(\tau_L)-\frac{a_0}{a_2}\left(\frac{a_1}{\varphi_1}+1\right)+y_d(t_0+\epsilon\tau_L). \tag{4.175}$$

The inner equations (4.173), (4.174) must be solved with the boundary conditions

$$X_L(0)=x_0^{\text{fb}}, \qquad Y_L(0)=y_0^{\text{fb}}. \tag{4.176}$$

To leading order in ϵ, $y_{\text{init}}^{\infty}(t_0+\epsilon\tau_L)$ simplifies to

$$y_{\text{init}}^{\infty}(t_0) = \int_{t_0}^{\infty} e^{(t_0-\tau)\varphi_1}\left(\frac{a_2 s_1}{s_2}x_d(\tau)-(a_1+\varphi_1)\,y_d(\tau)\right)d\tau \\ -\frac{1}{a_2}(\varphi_1+a_1)\,X_L(\tau_L)-\frac{a_0}{a_2}\left(\frac{a_1}{\varphi_1}+1\right)+y_d(t_0), \tag{4.177}$$

and the inner equations simplify to

$$\dot{X}_L(\tau_L)=0, \tag{4.178}$$
$$\dot{Y}_L(\tau_L)=\sqrt{s_2}\,(y_{\text{init}}^{\infty}(t_0)-Y_L(\tau_L))\,|b(x_L(\tau_L),Y_L(\tau_L))|.$$

The solution for X_L is

$$X_L(\tau_L)=x_0^{\text{fb}}, \tag{4.179}$$

and the equation for Y_L reduces to

$$\dot{Y}_L(\tau_L)=\sqrt{s_2}\,(y_{\text{init}}^{\infty}-Y_L(\tau_L))\,|b(x_0,Y_L(\tau_L))|, \tag{4.180}$$
$$Y_L(0)=y_0^{\text{fb}}, \tag{4.181}$$

with constant y_{init}^{∞} given by

$$y_{\text{init}}^{\infty} = \int_{t_0}^{\infty} e^{(t_0-\tau)\varphi_1}\left(\frac{a_2 s_1}{s_2}x_d(\tau)-(a_1+\varphi_1)\,y_d(\tau)\right)d\tau \\ -\frac{1}{a_2}(\varphi_1+a_1)\,x_0^{\text{fb}}-\frac{a_0}{a_2}\left(\frac{a_1}{\varphi_1}+1\right)+y_d(t_0). \tag{4.182}$$

Equation (4.180) is nonlinear and has no analytical solution in closed form. Remarkably, Eq. (4.180) has the same form as Eq. (4.65) for the inner boundary layer of open loop control. The initial boundary layers are governed by the same dynamics regardless of open or closed loop control. The only difference is the value

of the constant y_{init}^{∞}. Assuming that the coupling function $b(x, y)$ does not depend on y, Eq. (4.180) can immediately be solved,

$$Y_L(\tau_L) = \exp\left(-\sqrt{s_2}\tau_L \left|b(x_0)\right|\right)\left(y_0^{\text{fb}} - y_{\text{init}}^{\infty}\right) + y_{\text{init}}^{\infty}. \tag{4.183}$$

Finally, the matching procedure must be carried out. The matching conditions are

$$\lim_{\tau_L \to \infty} Y_L(\tau_L) = \lim_{t \to t_0} y_O(t), \qquad \lim_{\tau_L \to \infty} X_L(\tau_L) = \lim_{t \to t_0} x_O(t). \tag{4.184}$$

Equation (4.184) immediately yields

$$x_{\text{init}}^{\text{fb}} = x_0^{\text{fb}} \tag{4.185}$$

for the initial condition of the outer equation. The remaining matching condition for y is satisfied as well. The overlaps are obtained as

$$y_O(t_0) = y_{\text{init}}^{\infty}, \qquad x_O(t_0) = x_0^{\text{fb}}. \tag{4.186}$$

Inner and outer solutions are combined in a composite solution as

$$x_{\text{comp}}(t) = x_O(t) + X_L\left((t - t_0)/\epsilon\right) - x_0^{\text{fb}} = x_O(t), \tag{4.187}$$

$$\begin{aligned} y_{\text{comp}}(t) &= y_{\text{init}}^{\infty}(t) + Y_L\left((t - t_0)/\epsilon\right) - y_{\text{init}}^{\infty} \\ &= \int_t^{\infty} \left(e^{t\varphi_1} - e^{t_0\varphi_1}\right) e^{-\tau\varphi_1} \left(\frac{a_2 s_1}{s_2} x_d(\tau) - (a_1 + \varphi_1) y_d(\tau)\right) d\tau \\ &\quad - \int_{t_0}^{t} e^{(t_0 - \tau)\varphi_1} \left(\frac{a_2 s_1}{s_2} x_d(\tau) - (a_1 + \varphi_1) y_d(\tau)\right) d\tau \\ &\quad + \frac{1}{a_2}(\varphi_1 + a_1)\left(x_0^{\text{fb}} - x(t)\right) + y_d(t) - y_d(t_0) + Y_L\left((t - t_0)/\epsilon\right), \end{aligned} \tag{4.188}$$

Here, the outer solution $x_O(t)$ is given by Eq. (4.168), while the left inner solution $Y_L\left((t - t_0)/\epsilon\right)$ is given as the solution to Eq. (4.180).

Comparing the analytical result Eq. (4.158) with a numerical result requires the numerical solution of the Hamilton-Jacobi-Bellman equation (4.146). Unfortunately, this task is rather difficult and not pursued here. Instead, selected feedback-controlled state trajectories are compared with their open loop counterparts. An open loop control is determined for a specified value of the initial condition $\boldsymbol{x}_0$. Applying the same control signal to another initial condition usually fails. In contrast to that, a feedback-controlled state trajectory may start at an arbitrary initial condition $\boldsymbol{x}_0^{\text{fb}}$. If $\boldsymbol{x}_0 = \boldsymbol{x}_0^{\text{fb}}$, open loop and feedback-controlled state trajectories agree. Example 4.4 investigates the impact of selected initial conditions $\boldsymbol{x}_0 \neq \boldsymbol{x}_0^{\text{fb}}$ on the feedback-controlled state trajectory.

Example 4.4: Feedback-controlled FHN model

The feedback control Eq. (4.158) is applied to the activator-controlled FHN model, see Example 1.2 for details. The feedback-controlled state equations (4.159) and (4.160) are solved numerically. The desired trajectory and all parameters are the same as in Example 4.2, except for the amplitude $A_x = 5$ of the desired trajectory and an infinite terminal time $t_1 \to \infty$. Figure 4.11 shows the feedback-controlled state trajectory for inhibitor (left) and activator (right) for different values of the inhibitor initial condition x_0^{fb}. Figure 4.12 shows state trajectories for different values of the activator initial condition y_0^{fb}. In all plots, the blue solid line is the controlled state trajectory enforced by open loop control. During a transient, the feedback-controlled trajectories converge on the open loop-controlled state trajectory. However, the time scales of the transients are dramatically different. A deviation of the inhibitor initial condition decays very slowly, see Fig. 4.11. A deviation of the activator initial condition y_0^{fb} displays relaxation in form of a boundary layer. It decays on a time scale set by the boundary layer width ϵ, see Fig. 4.12.

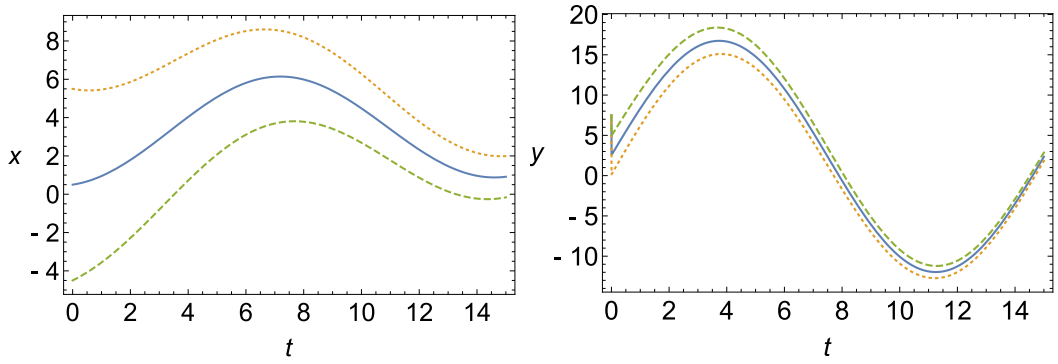

Figure 4.11.: Feedback-controlled state trajectories for different inhibitor initial conditions. The blue solid line with initial condition $\boldsymbol{x}_0^T = (x_0, y_0) = (0.5, 7.5)$ is the state trajectory enforced by open loop control. During a long transient, the feedback-controlled trajectories for $\left(x_0^{\text{fb}}, y_0^{\text{fb}}\right) = (x_0, y_0) + (5, 0)$ (orange dashed line) and $\left(x_0^{\text{fb}}, y_0^{\text{fb}}\right) = (x_0, y_0) - (5, 0)$ (green dashed line) converge on the blue solid trajectory.

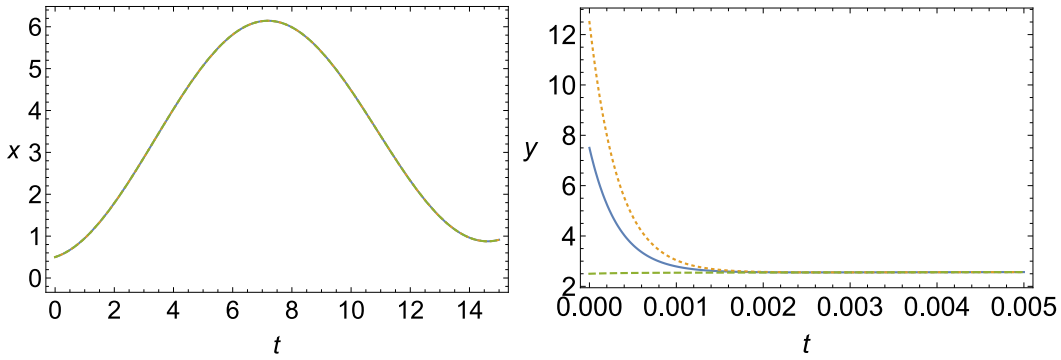

Figure 4.12.: Feedback-controlled state trajectories for different activator initial conditions. Compared with Fig. 4.11, the relaxation back on the state trajectory enforced by open loop control (blue solid line for $(x_0,\, y_0) = (0.5,\, 7.5)$) occurs much faster on a time scale set by ϵ. Feedback-controlled state trajectories are shown for $\left(x_0^{\mathrm{fb}},\, y_0^{\mathrm{fb}}\right) = (x_0,\, y_0) + (0,\, 5)$ and $\left(x_0^{\mathrm{fb}},\, y_0^{\mathrm{fb}}\right) = (x_0,\, y_0) - (0,\, 5)$ (green dashed line).

4.3.2. Continuous time-delayed feedback

Continuous time feedback cannot be applied to systems with fast dynamics. If measurement and processing of the system's state takes place on a time scale comparable with system dynamics, a delay is induced. The control signal fed back to the system cannot be assumed to depend on the current state $\boldsymbol{x}\,(t)$ of the system. Instead, the feedback signal depends on the delayed system state $\boldsymbol{x}\,(t-T)$. The time delay $T > 0$ accounts for the duration of measurement and information processing. Introducing artificial time delays and superpositions of continuous time and time-delayed feedback signals can be beneficial for stabilization by feedback. A prominent example is the stabilization of unstable periodic orbits in chaotic systems developed by Pyragas (Pyragas, 1992, 2006), see also the theme issue (Just et al., 2010).

The initial conditions x_0 and y_0 in Eqs. (4.94)-(4.96) Eq. (4.111) are assumed to depend on the delayed state components as $x_0 = x\,(t_0) = x\,(t-T)$ and $y_0 = y\,(t_0) = y\,(t-T)$, respectively. Every explicit appearance of t_0 in these equations is substituted by $t_0 \to t - T$. All constants which depend on time t_0, as e.g. x_{init}, become time dependent,

$$x_{\mathrm{init}} = x_{\mathrm{init}}\,(t)\,, \qquad y_{\mathrm{init}} = y_{\mathrm{init}}\,(t)\,, \qquad y_{\mathrm{end}} = y_{\mathrm{end}}\,(t)\,. \tag{4.189}$$

The outer solutions $x_O(t)$ and $y_O(t)$ given by Eqs. (4.33) and (4.34) become

$$\begin{aligned} x_O(t) = {} & \frac{a_1 a_2}{\varphi_1} \int_{t-T}^{t} y_d(\tau) \sinh(\varphi_1(t-\tau))\, d\tau + a_2 \int_{t-T}^{t} y_d(\tau) \cosh(\varphi_1(t-\tau))\, d\tau \\ & - \frac{a_2^2 s_1}{s_2 \varphi_1} \int_{t-T}^{t} x_d(\tau) \sinh(\varphi_1(t-\tau))\, d\tau + \frac{1}{\varphi_1} \sinh(T\varphi_1)(a_0 - a_2 y_d(t-T)) \\ & + \frac{a_0 a_1}{\varphi_1^2}(\cosh(T\varphi_1) - 1) + x_{\text{init}}(t)\left(\frac{a_1}{\varphi_1}\sinh(T\varphi_1) + \cosh(T\varphi_1)\right) \\ & + \frac{a_2}{\varphi_1} y_{\text{init}}(t) \sinh(T\varphi_1), \end{aligned} \tag{4.190}$$

and

$$\begin{aligned} y_O(t) = {} & \frac{a_0 a_2 s_1}{s_2 \varphi_1^2}(\cosh(T\varphi_1) - 1) + \frac{a_2 s_1}{s_2 \varphi_1} x_{\text{init}}(t) \sinh(T\varphi_1) \\ & + y_{\text{init}}(t) \cosh(T\varphi_1) - y_{\text{init}}(t) \frac{a_1}{\varphi_1} \sinh(T\varphi_1) \\ & - \frac{a_2 s_1}{s_2} \int_{t-T}^{t} x_d(\tau) \cosh(\varphi_1(t-\tau))\, d\tau + \frac{a_2 s_1 a_1}{s_2 \varphi_1} \int_{t-T}^{t} x_d(\tau) \sinh(\varphi_1(t-\tau))\, d\tau \\ & + \frac{a_2^2 s_1}{s_2 \varphi_1} \int_{t-T}^{t} y_d(\tau) \sinh(\varphi_1(t-\tau))\, d\tau \\ & + \frac{a_1}{\varphi_1} y_d(t-T) \sinh(T\varphi_1) + y_d(t) - y_d(t-T) \cosh(T\varphi_1). \end{aligned} \tag{4.191}$$

Due to the time delay, the integral terms in $x_O(t)$ and $y_O(t)$ do not vanish. The time-dependent constants $x_{\text{init}}(t)$, $y_{\text{init}}(t)$ are given by

$$x_{\text{init}}(t) = x(t-T), \tag{4.192}$$

and

$$\begin{aligned} y_{\text{init}}(t) = {} & \frac{s_2 \varphi_1}{\kappa(t)} \sinh((t-T-t_1)\varphi_1)\left(a_1 s_2 - a_2^2 \beta_1\right) y_d(t-T) \\ & + \frac{a_2 s_2 \varphi_1}{\kappa(t)} \sinh((t-T-t_1)\varphi_1)(a_0\beta_1 + x(t-T)(a_1\beta_1 + s_1)) \\ & - \frac{a_2 \beta_1}{\kappa(t)}\left(a_1^2 s_2 + a_2^2 s_1\right) \cosh((t_1 - t + T)\varphi_1)\, x(t-T) \\ & - \frac{a_2 a_0 s_2}{\kappa(t)}(a_1\beta_1 + s_1) \cosh((t_1 - t + T)\varphi_1) \\ & + \frac{s_2}{\kappa(t)}\left(a_1^2 s_2 + a_2^2 s_1\right) \cosh((t_1 - t + T)\varphi_1)\, y_d(t-T) \end{aligned}$$

$$+\frac{a_2\varphi_1 s_1}{\kappa(t)}\left(a_2^2\beta_1 - a_1 s_2\right)\int_{t-T}^{t_1} x_d(\tau)\sinh\left(\varphi_1(t_1-\tau)\right)\,d\tau$$

$$-\frac{a_2^2\varphi_1 s_2}{\kappa(t)}\left(a_1\beta_1 + s_1\right)\int_{t-T}^{t_1} y_d(\tau)\sinh\left(\varphi_1(t_1-\tau)\right)\,d\tau$$

$$-\frac{\beta_1 a_2^2\varphi_1^2 s_2}{\kappa(t)}\int_{t-T}^{t_1} y_d(\tau)\cosh\left(\varphi_1(t_1-\tau)\right)\,d\tau + \frac{a_2 a_0 s_2}{\kappa}\left(a_1\beta_1 + s_1\right)$$

$$+\frac{a_2\varphi_1^2 s_1 s_2}{\kappa(t)}\int_{t-T}^{t_1} x_d(\tau)\cosh\left(\varphi_1(t_1-\tau)\right)\,d\tau + \beta_1 x_1\frac{a_2}{\kappa}\left(a_1^2 s_2 + a_2^2 s_1\right). \quad (4.193)$$

The abbreviation $\kappa(t)$ is defined by

$$\kappa(t) = s_2\varphi_1\left(a_2^2\beta_1 - a_1 s_2\right)\sinh\left((t_1 - t + T)\varphi_1\right) + s_2^2\varphi_1^2\cosh\left((t_1 - t + T)\varphi_1\right). \quad (4.194)$$

Using Eqs. (4.190)-(4.194) together with the solution Eq. (4.111) for optimal open loop control, the optimal time-delayed feedback control signal $u_{\text{comp}}^{\text{fb}}(t)$ is obtained as

$$\begin{aligned} u_{\text{comp}}^{\text{fb}}(t) &= u_O(t) + U_L\left((t-t_0)/\epsilon\right) + U_R\left((t_1-t)/\epsilon\right) - u_O(t_0) - u_O(t_1) \\ &= u_O(t) + U_L(T/\epsilon) + U_R\left((t_1 - t + T)/\epsilon\right) - u_O(t-T) - u_O(t_1). \end{aligned} \quad (4.195)$$

Here, the outer control signal $u_O(t)$ is given by Eq. (4.101), and U_L and U_R are given by Eqs. (4.106) and (4.107), respectively. Caution has to be taken when evaluating an expression involving the time derivative $\dot{y}_O(t)$. The dot $\dot{y}_O(t)$ denotes the time derivative with respect to the current time t. It does not commute with the substitution $t_0 \to t - T$. Consequently, the time derivative has to be computed before the substitution $t_0 \to t - T$.

4.3.3. Discussion

A modification of the analytical results from Section 4.1 extends their scope to optimal feedback control. This requires knowledge about the dependency of the controlled state trajectory on its initial conditions. The essential idea is to replace the initial state $\boldsymbol{x}_0 = \boldsymbol{x}(t_0)$ with the monitored state $\boldsymbol{x}(t)$ of the controlled dynamical system. Numerically, optimal feedback is obtained by solving the Hamilton-Jacobi-Bellman equation (4.146).

The analytical approach yields a closed form expression for the continuous time feedback law Eq. (4.158). Remarkably, for infinite terminal time $t_1 \to \infty$, the

feedback-controlled state equation does not depend on the nonlinearity $R(x, y)$. The controlled state equations depend on ϵ and are singularly perturbed. Separating inner and outer equations reveals an initial boundary layer for the y-component. Its dynamics depends on the coupling function $b(x, y)$ and is identical in form to the boundary layers encountered for open loop control. For identical initial conditions, a feedback-controlled state trajectory is identical to a state trajectory enforced by open-loop control. For differing initial conditions, a feedback-controlled state trajectory relaxes onto the open loop trajectory. An initial deviation of the y-component converges swiftly. The relaxation is in form of a boundary layer with a time scale set by ϵ. Initial deviations of the x-component decay at low speed on a time scale independent of ϵ.

Stabilizing feedback control of unstable attractors received plentiful attention by the physics community, especially in the context of chaos control (Ott et al., 1990; Schöll and Schuster, 2007; Schimansky-Geier et al., 2007). However, optimality of these methods is rarely investigated. This is hardly surprising in view of the fundamental difficulties. Numerically solving the Hamilton-Jacobi-Bellman equation represents a difficult task itself. Analytical methods are largely restricted to linear systems which lack e.g. limit cycles and chaos. The analytical approach outlined in this section opens up a possibility to study the optimality of continuous time as well as time-delayed feedback stabilization of unstable attractors. An interesting option would be an optimal variant of the Pyragas control to stabilize periodic orbits in chaotic systems (Pyragas, 1992, 2006). Similarly, it is possible to investigate the optimality of techniques from mathematical control theory. An interesting problem concerns the optimality of feedback linearization. The continuous time feedback control given by Eq. (4.158) shares similarities with the control signal Eq. (1.35) obtained by feedback linearization. In both cases the control signal simply absorbs the nonlinearity $R(x, y)$. However, the optimal feedback law retains a nontrivial dependence on the coupling function $b(x, y)$ and results in nonlinear evolution equations for the feedback-controlled state.

4.4. General dynamical system

This section discusses optimal trajectory tracking for general dynamical systems. The task is to minimize the target functional

$$\begin{aligned}\mathcal{J}\left[\boldsymbol{x}(t), \boldsymbol{u}(t)\right] = &\int_{t_0}^{t_1} dt \frac{1}{2} \left(\boldsymbol{x}(t) - \boldsymbol{x}_d(t)\right)^T \boldsymbol{\mathcal{S}} \left(\boldsymbol{x}(t) - \boldsymbol{x}_d(t)\right) \\ &+ \frac{\epsilon^2}{2} \int_{t_0}^{t_1} dt \left(\boldsymbol{u}(t)\right)^2 \end{aligned} \tag{4.196}$$

subject to the dynamic constraints

$$\dot{\boldsymbol{x}}(t) = \boldsymbol{R}(\boldsymbol{x}(t)) + \boldsymbol{\mathcal{B}}(\boldsymbol{x}(t))\,\boldsymbol{u}(t), \qquad \boldsymbol{x}(t_0) = \boldsymbol{x}_0, \qquad \boldsymbol{x}(t_1) = \boldsymbol{x}_1. \tag{4.197}$$

Here, $\boldsymbol{\mathcal{S}} = \boldsymbol{\mathcal{S}}^T$ is a symmetric $n \times n$ matrix of weights. Only sharp terminal conditions $\boldsymbol{x}(t_1) = \boldsymbol{x}_1$ are discussed. The starting point for the perturbative treatment are the necessary optimality conditions,

$$\boldsymbol{0} = \epsilon^2 \boldsymbol{u}(t) + \boldsymbol{\mathcal{B}}^T(\boldsymbol{x}(t))\,\boldsymbol{\lambda}(t), \tag{4.198}$$

$$\dot{\boldsymbol{x}}(t) = \boldsymbol{R}(\boldsymbol{x}(t)) + \boldsymbol{\mathcal{B}}(\boldsymbol{x}(t))\,\boldsymbol{u}(t), \tag{4.199}$$

$$-\dot{\boldsymbol{\lambda}}(t) = \Big(\nabla \boldsymbol{R}^T(\boldsymbol{x}(t)) + \boldsymbol{u}^T(t)\,\nabla \boldsymbol{\mathcal{B}}^T(\boldsymbol{x}(t))\Big)\,\boldsymbol{\lambda}(t) + \boldsymbol{\mathcal{S}}(\boldsymbol{x}(t) - \boldsymbol{x}_d(t)), \tag{4.200}$$

together with the initial and terminal conditions

$$\boldsymbol{x}(t_0) = \boldsymbol{x}_0, \qquad \boldsymbol{x}(t_1) = \boldsymbol{x}_1. \tag{4.201}$$

See Section 3.1 for a derivation of Eqs. (4.198) and (4.200).

The idea of the analytical treatment is to utilize the two projectors $\boldsymbol{\mathcal{P}}_{\boldsymbol{\mathcal{S}}}(\boldsymbol{x})$ and $\boldsymbol{\mathcal{Q}}_{\boldsymbol{\mathcal{S}}}(\boldsymbol{x})$ defined by Eqs. (3.102) and (3.103) to split up the necessary optimality conditions. While the state projections $\boldsymbol{\mathcal{P}}_{\boldsymbol{\mathcal{S}}}(\boldsymbol{x})\,\boldsymbol{x}$ exhibit boundary layers, the state projections $\boldsymbol{\mathcal{Q}}_{\boldsymbol{\mathcal{S}}}(\boldsymbol{x})\,\boldsymbol{x}$ do not. The equations are rearranged to obtain a singularly perturbed system of differential equations. Inner and outer equations are determined by a perturbation expansion to leading order of the small parameter ϵ. A linearizing assumption similar to Section 2.3 results in linear outer equations which can be formally solved.

4.4.1. Rearranging the necessary optimality conditions

To shorten the notation, the time argument of $\boldsymbol{x}(t)$, $\boldsymbol{\lambda}(t)$ and $\boldsymbol{u}(t)$ is suppressed in this subsection and some abbreviating matrices are introduced. Let the $n \times n$ matrix $\boldsymbol{\Omega}_{\boldsymbol{\mathcal{S}}}(\boldsymbol{x})$ be defined by

$$\boldsymbol{\Omega}_{\boldsymbol{\mathcal{S}}}(\boldsymbol{x}) = \boldsymbol{\mathcal{B}}(\boldsymbol{x})\Big(\boldsymbol{\mathcal{B}}^T(\boldsymbol{x})\,\boldsymbol{\mathcal{S}}\boldsymbol{\mathcal{B}}(\boldsymbol{x})\Big)^{-1}\boldsymbol{\mathcal{B}}^T(\boldsymbol{x}). \tag{4.202}$$

A simple calculation shows that $\boldsymbol{\Omega}_{\boldsymbol{\mathcal{S}}}(\boldsymbol{x})$ is symmetric,

$$\begin{aligned}\boldsymbol{\Omega}_{\boldsymbol{\mathcal{S}}}^T(\boldsymbol{x}) &= \boldsymbol{\mathcal{B}}(\boldsymbol{x})\Big(\boldsymbol{\mathcal{B}}^T(\boldsymbol{x})\,\boldsymbol{\mathcal{S}}\boldsymbol{\mathcal{B}}(\boldsymbol{x})\Big)^{-T}\boldsymbol{\mathcal{B}}^T(\boldsymbol{x}) \\ &= \boldsymbol{\mathcal{B}}(\boldsymbol{x})\Big(\boldsymbol{\mathcal{B}}^T(\boldsymbol{x})\,\boldsymbol{\mathcal{S}}\boldsymbol{\mathcal{B}}(\boldsymbol{x})\Big)^{-1}\boldsymbol{\mathcal{B}}^T(\boldsymbol{x}) = \boldsymbol{\Omega}_{\boldsymbol{\mathcal{S}}}(\boldsymbol{x}).\end{aligned} \tag{4.203}$$

Note that $\boldsymbol{\mathcal{B}}^T(\boldsymbol{x})\,\boldsymbol{\mathcal{S}}\boldsymbol{\mathcal{B}}(\boldsymbol{x}) = \Big(\boldsymbol{\mathcal{B}}^T(\boldsymbol{x})\,\boldsymbol{\mathcal{S}}\boldsymbol{\mathcal{B}}(\boldsymbol{x})\Big)^T$ is a symmetric $p \times p$ matrix because $\boldsymbol{\mathcal{S}}$ is symmetric by assumption, and the inverse of a symmetric matrix is symmetric. Let the two $n \times n$ projectors $\boldsymbol{\mathcal{P}}_{\boldsymbol{\mathcal{S}}}(\boldsymbol{x})$ and $\boldsymbol{\mathcal{Q}}_{\boldsymbol{\mathcal{S}}}(\boldsymbol{x})$ be defined by

$$\boldsymbol{\mathcal{P}}_{\boldsymbol{\mathcal{S}}}(\boldsymbol{x}) = \boldsymbol{\Omega}_{\boldsymbol{\mathcal{S}}}(\boldsymbol{x})\,\boldsymbol{\mathcal{S}}, \qquad \boldsymbol{\mathcal{Q}}_{\boldsymbol{\mathcal{S}}}(\boldsymbol{x}) = \boldsymbol{1} - \boldsymbol{\mathcal{P}}_{\boldsymbol{\mathcal{S}}}(\boldsymbol{x}). \tag{4.204}$$

These projectors are derived during the discussion of singular optimal control in Section 3.4.2. $\boldsymbol{\mathcal{P}}_{\boldsymbol{\mathcal{S}}}(\boldsymbol{x})$ and $\boldsymbol{\mathcal{Q}}_{\boldsymbol{\mathcal{S}}}(\boldsymbol{x})$ are idempotent, $\boldsymbol{\mathcal{P}}_{\boldsymbol{\mathcal{S}}}^2(\boldsymbol{x}) = \boldsymbol{\mathcal{P}}_{\boldsymbol{\mathcal{S}}}(\boldsymbol{x})$ and $\boldsymbol{\mathcal{Q}}_{\boldsymbol{\mathcal{S}}}^2(\boldsymbol{x}) = \boldsymbol{\mathcal{Q}}_{\boldsymbol{\mathcal{S}}}(\boldsymbol{x})$. Furthermore, $\boldsymbol{\mathcal{P}}_{\boldsymbol{\mathcal{S}}}(\boldsymbol{x})$ and $\boldsymbol{\mathcal{Q}}_{\boldsymbol{\mathcal{S}}}(\boldsymbol{x})$ satisfy the relations

$$\boldsymbol{\mathcal{P}}_{\boldsymbol{\mathcal{S}}}(\boldsymbol{x})\,\boldsymbol{\mathcal{B}}(\boldsymbol{x}) = \boldsymbol{\mathcal{B}}(\boldsymbol{x}), \qquad \boldsymbol{\mathcal{Q}}_{\boldsymbol{\mathcal{S}}}(\boldsymbol{x})\,\boldsymbol{\mathcal{B}}(\boldsymbol{x}) = \mathbf{0}, \tag{4.205}$$

$$\boldsymbol{\mathcal{B}}^T(\boldsymbol{x})\,\boldsymbol{\mathcal{S}}\boldsymbol{\mathcal{P}}_{\boldsymbol{\mathcal{S}}}(\boldsymbol{x}) = \boldsymbol{\mathcal{B}}^T(\boldsymbol{x})\,\boldsymbol{\mathcal{S}}, \qquad \boldsymbol{\mathcal{B}}^T(\boldsymbol{x})\,\boldsymbol{\mathcal{S}}\boldsymbol{\mathcal{Q}}_{\boldsymbol{\mathcal{S}}} = \mathbf{0}. \tag{4.206}$$

Computing the transposed of $\boldsymbol{\mathcal{P}}_{\boldsymbol{\mathcal{S}}}(\boldsymbol{x})$ and $\boldsymbol{\mathcal{Q}}_{\boldsymbol{\mathcal{S}}}(\boldsymbol{x})$ yields

$$\boldsymbol{\mathcal{P}}_{\boldsymbol{\mathcal{S}}}^T(\boldsymbol{x}) = \boldsymbol{\mathcal{S}}^T\boldsymbol{\Omega}_{\boldsymbol{\mathcal{S}}}^T(\boldsymbol{x}) = \boldsymbol{\mathcal{S}}\boldsymbol{\Omega}_{\boldsymbol{\mathcal{S}}}(\boldsymbol{x}) \neq \boldsymbol{\mathcal{P}}_{\boldsymbol{\mathcal{S}}}(\boldsymbol{x}), \tag{4.207}$$

and analogously for $\boldsymbol{\mathcal{Q}}_{\boldsymbol{\mathcal{S}}}(\boldsymbol{x})$. Equation (4.207) shows that $\boldsymbol{\mathcal{P}}_{\boldsymbol{\mathcal{S}}}(\boldsymbol{x})$, and therefore also $\boldsymbol{\mathcal{Q}}_{\boldsymbol{\mathcal{S}}}(\boldsymbol{x})$, is not symmetric. However, $\boldsymbol{\mathcal{P}}_{\boldsymbol{\mathcal{S}}}^T(\boldsymbol{x})$ satisfies the convenient property

$$\boldsymbol{\mathcal{P}}_{\boldsymbol{\mathcal{S}}}^T(\boldsymbol{x})\,\boldsymbol{\mathcal{S}} = \boldsymbol{\mathcal{S}}\boldsymbol{\Omega}_{\boldsymbol{\mathcal{S}}}(\boldsymbol{x})\,\boldsymbol{\mathcal{S}} = \boldsymbol{\mathcal{S}}\boldsymbol{\mathcal{P}}_{\boldsymbol{\mathcal{S}}}(\boldsymbol{x}), \tag{4.208}$$

which implies

$$\boldsymbol{\mathcal{P}}_{\boldsymbol{\mathcal{S}}}^T(\boldsymbol{x})\,\boldsymbol{\mathcal{S}} = \boldsymbol{\mathcal{P}}_{\boldsymbol{\mathcal{S}}}^T(\boldsymbol{x})\,\boldsymbol{\mathcal{P}}_{\boldsymbol{\mathcal{S}}}^T(\boldsymbol{x})\,\boldsymbol{\mathcal{S}} = \boldsymbol{\mathcal{P}}_{\boldsymbol{\mathcal{S}}}^T(\boldsymbol{x})\,\boldsymbol{\mathcal{S}}\boldsymbol{\mathcal{P}}_{\boldsymbol{\mathcal{S}}}(\boldsymbol{x}), \tag{4.209}$$

and similarly for $\boldsymbol{\mathcal{S}}\boldsymbol{\mathcal{Q}}_{\boldsymbol{\mathcal{S}}}(\boldsymbol{x})$. The product of $\boldsymbol{\Omega}_{\boldsymbol{\mathcal{S}}}(\boldsymbol{x})$ with $\boldsymbol{\mathcal{P}}_{\boldsymbol{\mathcal{S}}}^T(\boldsymbol{x})$ yields

$$\begin{aligned}\boldsymbol{\Omega}_{\boldsymbol{\mathcal{S}}}(\boldsymbol{x})\,\boldsymbol{\mathcal{P}}_{\boldsymbol{\mathcal{S}}}^T(\boldsymbol{x}) &= \boldsymbol{\Omega}_{\boldsymbol{\mathcal{S}}}(\boldsymbol{x})\,\boldsymbol{\mathcal{S}}\boldsymbol{\Omega}_{\boldsymbol{\mathcal{S}}}(\boldsymbol{x}) \\ &= \boldsymbol{\mathcal{B}}(\boldsymbol{x})\left(\boldsymbol{\mathcal{B}}^T(\boldsymbol{x})\,\boldsymbol{\mathcal{S}}\boldsymbol{\mathcal{B}}(\boldsymbol{x})\right)^{-1}\boldsymbol{\mathcal{B}}^T(\boldsymbol{x}) = \boldsymbol{\Omega}_{\boldsymbol{\mathcal{S}}}(\boldsymbol{x}),\end{aligned} \tag{4.210}$$

and

$$\boldsymbol{\Omega}_{\boldsymbol{\mathcal{S}}}(\boldsymbol{x})\,\boldsymbol{\mathcal{P}}_{\boldsymbol{\mathcal{S}}}^T(\boldsymbol{x})\,\boldsymbol{\mathcal{S}} = \boldsymbol{\Omega}_{\boldsymbol{\mathcal{S}}}(\boldsymbol{x})\,\boldsymbol{\mathcal{S}} = \boldsymbol{\mathcal{P}}_{\boldsymbol{\mathcal{S}}}(\boldsymbol{x}). \tag{4.211}$$

Let the $n \times n$ matrix $\boldsymbol{\Gamma}_{\boldsymbol{\mathcal{S}}}(\boldsymbol{x})$ be defined by

$$\boldsymbol{\Gamma}_{\boldsymbol{\mathcal{S}}}(\boldsymbol{x}) = \boldsymbol{\mathcal{S}}\boldsymbol{\mathcal{B}}(\boldsymbol{x})\left(\boldsymbol{\mathcal{B}}^T(\boldsymbol{x})\,\boldsymbol{\mathcal{S}}\boldsymbol{\mathcal{B}}(\boldsymbol{x})\right)^{-2}\boldsymbol{\mathcal{B}}^T(\boldsymbol{x})\,\boldsymbol{\mathcal{S}}. \tag{4.212}$$

$\boldsymbol{\Gamma}_{\boldsymbol{\mathcal{S}}}(\boldsymbol{x})$ is symmetric,

$$\begin{aligned}\boldsymbol{\Gamma}_{\boldsymbol{\mathcal{S}}}^T(\boldsymbol{x}) &= \left(\boldsymbol{\mathcal{S}}\boldsymbol{\mathcal{B}}(\boldsymbol{x})\left(\boldsymbol{\mathcal{B}}^T(\boldsymbol{x})\,\boldsymbol{\mathcal{S}}\boldsymbol{\mathcal{B}}(\boldsymbol{x})\right)^{-2}\boldsymbol{\mathcal{B}}^T(\boldsymbol{x})\,\boldsymbol{\mathcal{S}}\right)^T \\ &= \boldsymbol{\mathcal{S}}\boldsymbol{\mathcal{B}}(\boldsymbol{x})\left(\boldsymbol{\mathcal{B}}^T(\boldsymbol{x})\,\boldsymbol{\mathcal{S}}\boldsymbol{\mathcal{B}}(\boldsymbol{x})\right)^{-2T}\boldsymbol{\mathcal{B}}^T(\boldsymbol{x})\,\boldsymbol{\mathcal{S}} = \boldsymbol{\Gamma}_{\boldsymbol{\mathcal{S}}}(\boldsymbol{x}),\end{aligned} \tag{4.213}$$

and satisfies

$$\begin{aligned}&\boldsymbol{\Gamma}_{\boldsymbol{\mathcal{S}}}(\boldsymbol{x})\,\boldsymbol{\mathcal{P}}_{\boldsymbol{\mathcal{S}}}(\boldsymbol{x}) \\ =&\boldsymbol{\mathcal{S}}\boldsymbol{\mathcal{B}}(\boldsymbol{x})\left(\boldsymbol{\mathcal{B}}^T(\boldsymbol{x})\,\boldsymbol{\mathcal{S}}\boldsymbol{\mathcal{B}}(\boldsymbol{x})\right)^{-2}\boldsymbol{\mathcal{B}}^T(\boldsymbol{x})\,\boldsymbol{\mathcal{S}}\boldsymbol{\mathcal{B}}(\boldsymbol{x})\left(\boldsymbol{\mathcal{B}}^T(\boldsymbol{x})\,\boldsymbol{\mathcal{S}}\boldsymbol{\mathcal{B}}(\boldsymbol{x})\right)^{-1}\boldsymbol{\mathcal{B}}^T(\boldsymbol{x})\,\boldsymbol{\mathcal{S}} \\ =&\boldsymbol{\mathcal{S}}\boldsymbol{\mathcal{B}}(\boldsymbol{x})\left(\boldsymbol{\mathcal{B}}^T(\boldsymbol{x})\,\boldsymbol{\mathcal{S}}\boldsymbol{\mathcal{B}}(\boldsymbol{x})\right)^{-2}\boldsymbol{\mathcal{B}}^T(\boldsymbol{x})\,\boldsymbol{\mathcal{S}} = \boldsymbol{\Gamma}_{\boldsymbol{\mathcal{S}}}(\boldsymbol{x}).\end{aligned} \tag{4.214}$$

Transposing yields

$$\left(\boldsymbol{\Gamma}_{\boldsymbol{\mathcal{S}}}\left(\boldsymbol{x}\right)\boldsymbol{\mathcal{P}}_{\boldsymbol{\mathcal{S}}}\left(\boldsymbol{x}\right)\right)^{T}=\boldsymbol{\mathcal{P}}_{\boldsymbol{\mathcal{S}}}^{T}\left(\boldsymbol{x}\right)\boldsymbol{\Gamma}_{\boldsymbol{\mathcal{S}}}^{T}\left(\boldsymbol{x}\right)=\boldsymbol{\mathcal{P}}_{\boldsymbol{\mathcal{S}}}^{T}\left(\boldsymbol{x}\right)\boldsymbol{\Gamma}_{\boldsymbol{\mathcal{S}}}\left(\boldsymbol{x}\right)=\boldsymbol{\Gamma}_{\boldsymbol{\mathcal{S}}}\left(\boldsymbol{x}\right). \tag{4.215}$$

The projectors $\boldsymbol{\mathcal{P}}_{\boldsymbol{\mathcal{S}}}\left(\boldsymbol{x}\right)$ and $\boldsymbol{\mathcal{Q}}_{\boldsymbol{\mathcal{S}}}\left(\boldsymbol{x}\right)$ are used to partition the state $\boldsymbol{x}$,

$$\boldsymbol{x}=\boldsymbol{\mathcal{P}}_{\boldsymbol{\mathcal{S}}}\left(\boldsymbol{x}\right)\boldsymbol{x}+\boldsymbol{\mathcal{Q}}_{\boldsymbol{\mathcal{S}}}\left(\boldsymbol{x}\right)\boldsymbol{x}. \tag{4.216}$$

The controlled state equation (4.199) is split in two parts,

$$\boldsymbol{\mathcal{P}}_{\boldsymbol{\mathcal{S}}}\left(\boldsymbol{x}\right)\dot{\boldsymbol{x}}=\boldsymbol{\mathcal{P}}_{\boldsymbol{\mathcal{S}}}\left(\boldsymbol{x}\right)\boldsymbol{R}\left(\boldsymbol{x}\right)+\boldsymbol{\mathcal{B}}\left(\boldsymbol{x}\right)\boldsymbol{u}, \tag{4.217}$$

$$\boldsymbol{\mathcal{Q}}_{\boldsymbol{\mathcal{S}}}\left(\boldsymbol{x}\right)\dot{\boldsymbol{x}}=\boldsymbol{\mathcal{Q}}_{\boldsymbol{\mathcal{S}}}\left(\boldsymbol{x}\right)\boldsymbol{R}\left(\boldsymbol{x}\right). \tag{4.218}$$

The initial and terminal conditions are split up as well,

$$\boldsymbol{\mathcal{P}}_{\boldsymbol{\mathcal{S}}}\left(\boldsymbol{x}\left(t_{0}\right)\right)\boldsymbol{x}\left(t_{0}\right)=\boldsymbol{\mathcal{P}}_{\boldsymbol{\mathcal{S}}}\left(\boldsymbol{x}_{0}\right)\boldsymbol{x}_{0},\qquad\boldsymbol{\mathcal{Q}}_{\boldsymbol{\mathcal{S}}}\left(\boldsymbol{x}\left(t_{0}\right)\right)\boldsymbol{x}\left(t_{0}\right)=\boldsymbol{\mathcal{Q}}_{\boldsymbol{\mathcal{S}}}\left(\boldsymbol{x}_{0}\right)\boldsymbol{x}_{0}, \tag{4.219}$$

$$\boldsymbol{\mathcal{P}}_{\boldsymbol{\mathcal{S}}}\left(\boldsymbol{x}\left(t_{1}\right)\right)\boldsymbol{x}\left(t_{1}\right)=\boldsymbol{\mathcal{P}}_{\boldsymbol{\mathcal{S}}}\left(\boldsymbol{x}_{1}\right)\boldsymbol{x}_{1},\qquad\boldsymbol{\mathcal{Q}}_{\boldsymbol{\mathcal{S}}}\left(\boldsymbol{x}\left(t_{1}\right)\right)\boldsymbol{x}\left(t_{1}\right)=\boldsymbol{\mathcal{Q}}_{\boldsymbol{\mathcal{S}}}\left(\boldsymbol{x}_{1}\right)\boldsymbol{x}_{1}. \tag{4.220}$$

With the help of the relation $\boldsymbol{\mathcal{B}}^{T}\left(\boldsymbol{x}\right)\boldsymbol{\mathcal{S}}\boldsymbol{\mathcal{P}}_{\boldsymbol{\mathcal{S}}}\left(\boldsymbol{x}\right)=\boldsymbol{\mathcal{B}}^{T}\left(\boldsymbol{x}\right)\boldsymbol{\mathcal{S}}$, Eq. (4.217) is solved to obtain an expression for the control $\boldsymbol{u}$ in terms of the controlled state trajectory $\boldsymbol{x}$,

$$\begin{aligned}\boldsymbol{u}&=\left(\boldsymbol{\mathcal{B}}^{T}\left(\boldsymbol{x}\right)\boldsymbol{\mathcal{S}}\boldsymbol{\mathcal{B}}\left(\boldsymbol{x}\right)\right)^{-1}\boldsymbol{\mathcal{B}}^{T}\left(\boldsymbol{x}\right)\boldsymbol{\mathcal{S}}\left(\dot{\boldsymbol{x}}-\boldsymbol{R}\left(\boldsymbol{x}\right)\right)\\&=\boldsymbol{\mathcal{B}}_{\boldsymbol{\mathcal{S}}}^{g}\left(\boldsymbol{x}\right)\left(\dot{\boldsymbol{x}}-\boldsymbol{R}\left(\boldsymbol{x}\right)\right).\end{aligned} \tag{4.221}$$

The $p\times n$ matrix $\boldsymbol{\mathcal{B}}_{\boldsymbol{\mathcal{S}}}^{g}\left(\boldsymbol{x}\right)$ is a generalized reflexive inverse of $\boldsymbol{\mathcal{B}}\left(\boldsymbol{x}\right)$, see Appendix A.2.1, and defined by

$$\boldsymbol{\mathcal{B}}_{\boldsymbol{\mathcal{S}}}^{g}\left(\boldsymbol{x}\right)=\left(\boldsymbol{\mathcal{B}}^{T}\left(\boldsymbol{x}\right)\boldsymbol{\mathcal{S}}\boldsymbol{\mathcal{B}}\left(\boldsymbol{x}\right)\right)^{-1}\boldsymbol{\mathcal{B}}^{T}\left(\boldsymbol{x}\right)\boldsymbol{\mathcal{S}}. \tag{4.222}$$

The matrix $\boldsymbol{\mathcal{B}}_{\boldsymbol{\mathcal{S}}}^{g}\left(\boldsymbol{x}\right)$ can be used to rewrite the matrices $\boldsymbol{\mathcal{P}}_{\boldsymbol{\mathcal{S}}}\left(\boldsymbol{x}\right)$ and $\boldsymbol{\Gamma}_{\boldsymbol{\mathcal{S}}}\left(\boldsymbol{x}\right)$ as

$$\boldsymbol{\mathcal{P}}_{\boldsymbol{\mathcal{S}}}\left(\boldsymbol{x}\right)=\boldsymbol{\mathcal{B}}\left(\boldsymbol{x}\right)\boldsymbol{\mathcal{B}}_{\boldsymbol{\mathcal{S}}}^{g}\left(\boldsymbol{x}\right),\qquad\boldsymbol{\Gamma}_{\boldsymbol{\mathcal{S}}}\left(\boldsymbol{x}\right)=\boldsymbol{\mathcal{B}}_{\boldsymbol{\mathcal{S}}}^{gT}\left(\boldsymbol{x}\right)\boldsymbol{\mathcal{B}}_{\boldsymbol{\mathcal{S}}}^{g}\left(\boldsymbol{x}\right), \tag{4.223}$$

respectively. The $n\times p$ matrix $\boldsymbol{\mathcal{B}}_{\boldsymbol{\mathcal{S}}}^{gT}\left(\boldsymbol{x}\right)$ is the transposed of $\boldsymbol{\mathcal{B}}_{\boldsymbol{\mathcal{S}}}^{g}\left(\boldsymbol{x}\right)$.

The solution for $\boldsymbol{u}$ is inserted in the stationarity condition Eq. (4.198) to yield

$$\begin{aligned}\boldsymbol{0}&=\epsilon^{2}\boldsymbol{u}^{T}+\boldsymbol{\lambda}^{T}\boldsymbol{\mathcal{B}}\left(\boldsymbol{x}\right)\\&=\epsilon^{2}\left(\dot{\boldsymbol{x}}^{T}-\boldsymbol{R}^{T}\left(\boldsymbol{x}\right)\right)\boldsymbol{\mathcal{B}}_{\boldsymbol{\mathcal{S}}}^{gT}\left(\boldsymbol{x}\right)+\boldsymbol{\lambda}^{T}\boldsymbol{\mathcal{B}}\left(\boldsymbol{x}\right).\end{aligned} \tag{4.224}$$

Equation (4.224) is utilized to eliminate any occurrence of the part $\boldsymbol{\mathcal{P}}_{\boldsymbol{\mathcal{S}}}^{T}\left(\boldsymbol{x}\right)\boldsymbol{\lambda}$ in all equations. In contrast to the state $\boldsymbol{x}$, cf. Eq. (4.216), the co-state is split up with the transposed projectors $\boldsymbol{\mathcal{P}}_{\boldsymbol{\mathcal{S}}}^{T}\left(\boldsymbol{x}\right)$ and $\boldsymbol{\mathcal{Q}}_{\boldsymbol{\mathcal{S}}}^{T}\left(\boldsymbol{x}\right)$,

$$\boldsymbol{\lambda}=\boldsymbol{\mathcal{P}}_{\boldsymbol{\mathcal{S}}}^{T}\left(\boldsymbol{x}\right)\boldsymbol{\lambda}+\boldsymbol{\mathcal{Q}}_{\boldsymbol{\mathcal{S}}}^{T}\left(\boldsymbol{x}\right)\boldsymbol{\lambda}. \tag{4.225}$$

Multiplying Eq. (4.224) with $\boldsymbol{\mathcal{B}}^g_{\boldsymbol{\mathcal{S}}}(\boldsymbol{x})$ from the right and using Eq. (4.223) yields an expression for $\boldsymbol{\mathcal{P}}^T_{\boldsymbol{\mathcal{S}}}(\boldsymbol{x})\,\boldsymbol{\lambda}$,

$$\mathbf{0} = \epsilon^2 \left(\dot{\boldsymbol{x}}^T - \boldsymbol{R}^T(\boldsymbol{x})\right) \boldsymbol{\Gamma}_{\boldsymbol{\mathcal{S}}}(\boldsymbol{x}) + \boldsymbol{\lambda}^T \boldsymbol{\mathcal{P}}_{\boldsymbol{\mathcal{S}}}(\boldsymbol{x}). \tag{4.226}$$

Transposing the last equation and exploiting the symmetry of $\boldsymbol{\Gamma}_{\boldsymbol{\mathcal{S}}}(\boldsymbol{x})$, Eq. (4.213), yields

$$\boldsymbol{\mathcal{P}}^T_{\boldsymbol{\mathcal{S}}}(\boldsymbol{x})\,\boldsymbol{\lambda} = -\epsilon^2 \boldsymbol{\Gamma}_{\boldsymbol{\mathcal{S}}}(\boldsymbol{x}) \left(\dot{\boldsymbol{x}} - \boldsymbol{R}(\boldsymbol{x})\right). \tag{4.227}$$

Equation (4.227) is valid for all times $t_0 \leq t \leq t_1$. Applying the time derivative gives

$$\begin{aligned}\mathbf{0} = {} & \epsilon^2 \boldsymbol{\Gamma}_{\boldsymbol{\mathcal{S}}}(\boldsymbol{x}) \left(\ddot{\boldsymbol{x}} - \nabla \boldsymbol{R}(\boldsymbol{x})\,\dot{\boldsymbol{x}}\right) + \epsilon^2 \dot{\boldsymbol{\Gamma}}_{\boldsymbol{\mathcal{S}}}(\boldsymbol{x}) \left(\dot{\boldsymbol{x}} - \boldsymbol{R}(\boldsymbol{x})\right) \\ & + \dot{\boldsymbol{\mathcal{P}}}^T_{\boldsymbol{\mathcal{S}}}(\boldsymbol{x})\,\boldsymbol{\lambda} + \boldsymbol{\mathcal{P}}^T_{\boldsymbol{\mathcal{S}}}(\boldsymbol{x})\,\dot{\boldsymbol{\lambda}}.\end{aligned} \tag{4.228}$$

The short hand notations

$$\dot{\boldsymbol{\Gamma}}_{\boldsymbol{\mathcal{S}}}(\boldsymbol{x}) = \nabla \boldsymbol{\Gamma}_{\boldsymbol{\mathcal{S}}}(\boldsymbol{x})\,\dot{\boldsymbol{x}}, \qquad \dot{\boldsymbol{\mathcal{P}}}^T_{\boldsymbol{\mathcal{S}}}(\boldsymbol{x}) = \left(\dot{\boldsymbol{x}}^T \nabla\right) \boldsymbol{\mathcal{P}}^T_{\boldsymbol{\mathcal{S}}}(\boldsymbol{x}), \tag{4.229}$$

were introduced in Eq. (4.228). Splitting the co-state $\boldsymbol{\lambda}$ as in Eq. (4.225) and using Eq. (4.227) to eliminate $\boldsymbol{\mathcal{P}}^T_{\boldsymbol{\mathcal{S}}}(\boldsymbol{x})\,\boldsymbol{\lambda}$ leads to

$$\begin{aligned}-\boldsymbol{\mathcal{P}}^T_{\boldsymbol{\mathcal{S}}}(\boldsymbol{x})\,\dot{\boldsymbol{\lambda}} = {} & \epsilon^2 \boldsymbol{\Gamma}_{\boldsymbol{\mathcal{S}}}(\boldsymbol{x}) \left(\ddot{\boldsymbol{x}} - \nabla \boldsymbol{R}(\boldsymbol{x})\,\dot{\boldsymbol{x}}\right) + \dot{\boldsymbol{\mathcal{P}}}^T_{\boldsymbol{\mathcal{S}}}(\boldsymbol{x})\,\boldsymbol{\mathcal{Q}}^T_{\boldsymbol{\mathcal{S}}}(\boldsymbol{x})\,\boldsymbol{\lambda} \\ & + \epsilon^2 \left(\dot{\boldsymbol{\Gamma}}_{\boldsymbol{\mathcal{S}}}(\boldsymbol{x}) - \dot{\boldsymbol{\mathcal{P}}}^T_{\boldsymbol{\mathcal{S}}}(\boldsymbol{x})\,\boldsymbol{\Gamma}_{\boldsymbol{\mathcal{S}}}(\boldsymbol{x})\right) \left(\dot{\boldsymbol{x}} - \boldsymbol{R}(\boldsymbol{x})\right).\end{aligned} \tag{4.230}$$

Equation is an expression for $\boldsymbol{\mathcal{P}}^T_{\boldsymbol{\mathcal{S}}}(\boldsymbol{x})\,\dot{\boldsymbol{\lambda}}$ independent of $\boldsymbol{\mathcal{P}}^T_{\boldsymbol{\mathcal{S}}}(\boldsymbol{x})\,\boldsymbol{\lambda}$.

A similar procedure is performed for the adjoint equation (4.200). Eliminating the control signal $\boldsymbol{u}$ from Eq. (4.200) gives

$$-\dot{\boldsymbol{\lambda}} = \left(\nabla \boldsymbol{R}^T(\boldsymbol{x}) + \left(\dot{\boldsymbol{x}}^T - \boldsymbol{R}^T(\boldsymbol{x})\right) \boldsymbol{\mathcal{B}}^{gT}_{\boldsymbol{\mathcal{S}}}(\boldsymbol{x})\,\nabla \boldsymbol{\mathcal{B}}^T(\boldsymbol{x})\right) \boldsymbol{\lambda} + \boldsymbol{\mathcal{S}}\left(\boldsymbol{x} - \boldsymbol{x}_d(t)\right). \tag{4.231}$$

The expression $\boldsymbol{\mathcal{B}}^{gT}_{\boldsymbol{\mathcal{S}}}(\boldsymbol{x})\,\nabla \boldsymbol{\mathcal{B}}^T(\boldsymbol{x})$ is a third order tensor with $n \times n \times n$ components defined as (see also Eq. (3.17) for the meaning of $\nabla \boldsymbol{\mathcal{B}}(\boldsymbol{x})$)

$$\left(\nabla \boldsymbol{\mathcal{B}}(\boldsymbol{x})\,\boldsymbol{\mathcal{B}}^g_{\boldsymbol{\mathcal{S}}}(\boldsymbol{x})\right)_{ijk} = \sum_{l=1}^{p} \frac{\partial}{\partial x_j} \mathcal{B}_{il}(\boldsymbol{x})\,\mathcal{B}^g_{\boldsymbol{\mathcal{S}},lk}(\boldsymbol{x}). \tag{4.232}$$

The product $\boldsymbol{\lambda}^T \nabla \boldsymbol{\mathcal{B}}(\boldsymbol{x})\,\boldsymbol{\mathcal{B}}^g_{\boldsymbol{\mathcal{S}}}(\boldsymbol{x})\,\boldsymbol{x}$ yields an n-component row vector with entries

$$\left(\boldsymbol{\lambda}^T \nabla \boldsymbol{\mathcal{B}}(\boldsymbol{x})\,\boldsymbol{\mathcal{B}}^g_{\boldsymbol{\mathcal{S}}}(\boldsymbol{x})\,\boldsymbol{x}\right)_j = \sum_{i=1}^{n} \sum_{k=1}^{n} \sum_{l=1}^{p} \lambda_i \frac{\partial}{\partial x_j} \mathcal{B}_{il}(\boldsymbol{x})\,\mathcal{B}^g_{\boldsymbol{\mathcal{S}},lk}(\boldsymbol{x})\,x_k. \tag{4.233}$$

Tho shorten the notation, the $n \times n$ matrix

$$\boldsymbol{\mathcal{W}}(\boldsymbol{x}, \boldsymbol{y}) = \nabla \boldsymbol{\mathcal{B}}(\boldsymbol{x})\,\boldsymbol{\mathcal{B}}^g_{\boldsymbol{\mathcal{S}}}(\boldsymbol{x}) \left(\boldsymbol{y} - \boldsymbol{R}(\boldsymbol{x})\right) \tag{4.234}$$

with entries

$$\mathcal{W}_{ij}(\boldsymbol{x},\boldsymbol{y}) = \sum_{k=1}^{n}\sum_{l=1}^{p}\frac{\partial}{\partial x_j}\mathcal{B}_{il}(\boldsymbol{x})\,\mathcal{B}^g_{\boldsymbol{\mathcal{S}},lk}(\boldsymbol{x})\,(y_k - R_k(\boldsymbol{x})) \tag{4.235}$$

is introduced. Using Eq. (4.223) yields the identity

$$\frac{\partial}{\partial x_j}\boldsymbol{\mathcal{B}}(\boldsymbol{x})\,\boldsymbol{\mathcal{B}}^g_{\boldsymbol{\mathcal{S}}}(\boldsymbol{x}) = \frac{\partial}{\partial x_j}\boldsymbol{\mathcal{P}}_{\boldsymbol{\mathcal{S}}}(\boldsymbol{x}) - \boldsymbol{\mathcal{B}}(\boldsymbol{x})\,\frac{\partial}{\partial x_j}\boldsymbol{\mathcal{B}}^g_{\boldsymbol{\mathcal{S}}}(\boldsymbol{x}), \tag{4.236}$$

and the entries of $\boldsymbol{\mathcal{W}}(\boldsymbol{x},\boldsymbol{y})$ can be expressed as

$$\begin{aligned}\mathcal{W}_{ij}(\boldsymbol{x},\boldsymbol{y}) = &\sum_{k=1}^{n}\frac{\partial}{\partial x_j}\mathcal{P}_{\boldsymbol{\mathcal{S}},ik}(\boldsymbol{x})\,(y_k - R_k(\boldsymbol{x}))\\ &-\sum_{k=1}^{n}\sum_{l=1}^{p}\mathcal{B}_{il}(\boldsymbol{x})\,\frac{\partial}{\partial x_j}\mathcal{B}^g_{\boldsymbol{\mathcal{S}},lk}(\boldsymbol{x})\,(y_k - R_k(\boldsymbol{x})).\end{aligned} \tag{4.237}$$

Because of $\boldsymbol{\mathcal{Q}}_{\boldsymbol{\mathcal{S}}}(\boldsymbol{x})\,\boldsymbol{\mathcal{B}}(\boldsymbol{x}) = \mathbf{0}$, the product $\boldsymbol{\mathcal{Q}}_{\boldsymbol{\mathcal{S}}}(\boldsymbol{x})\,\boldsymbol{\mathcal{W}}(\boldsymbol{x},\boldsymbol{y})$ is

$$\sum_{i=1}^{n}\mathcal{Q}_{\boldsymbol{\mathcal{S}},li}(\boldsymbol{x})\,\mathcal{W}_{ij}(\boldsymbol{x},\boldsymbol{y}) = \sum_{i=1}^{n}\sum_{k=1}^{n}\mathcal{Q}_{\boldsymbol{\mathcal{S}},li}(\boldsymbol{x})\,\frac{\partial}{\partial x_j}\mathcal{P}_{\boldsymbol{\mathcal{S}},ik}(\boldsymbol{x})\,(y_k - R_k(\boldsymbol{x})). \tag{4.238}$$

In terms of the matrix $\boldsymbol{\mathcal{W}}(\boldsymbol{x},\boldsymbol{y})$, Eq. (4.231) assumes the shorter form

$$-\dot{\boldsymbol{\lambda}} = \left(\nabla\boldsymbol{R}^T(\boldsymbol{x}) + \boldsymbol{\mathcal{W}}^T(\boldsymbol{x},\dot{\boldsymbol{x}})\right)\boldsymbol{\lambda} + \boldsymbol{\mathcal{S}}\,(\boldsymbol{x} - \boldsymbol{x}_d(t)). \tag{4.239}$$

With the help of the projectors $\boldsymbol{\mathcal{P}}^T_{\boldsymbol{\mathcal{S}}}$ and $\boldsymbol{\mathcal{Q}}^T_{\boldsymbol{\mathcal{S}}}$, Eq. (4.239) is split up in two parts,

$$-\boldsymbol{\mathcal{P}}^T_{\boldsymbol{\mathcal{S}}}(\boldsymbol{x})\,\dot{\boldsymbol{\lambda}} = \boldsymbol{\mathcal{P}}^T_{\boldsymbol{\mathcal{S}}}(\boldsymbol{x})\left(\nabla\boldsymbol{R}^T(\boldsymbol{x}) + \boldsymbol{\mathcal{W}}^T(\boldsymbol{x},\dot{\boldsymbol{x}})\right)\boldsymbol{\lambda} + \boldsymbol{\mathcal{P}}^T_{\boldsymbol{\mathcal{S}}}(\boldsymbol{x})\,\boldsymbol{\mathcal{S}}\,(\boldsymbol{x} - \boldsymbol{x}_d(t)), \tag{4.240}$$

$$-\boldsymbol{\mathcal{Q}}^T_{\boldsymbol{\mathcal{S}}}(\boldsymbol{x})\,\dot{\boldsymbol{\lambda}} = \boldsymbol{\mathcal{Q}}^T_{\boldsymbol{\mathcal{S}}}(\boldsymbol{x})\left(\nabla\boldsymbol{R}^T(\boldsymbol{x}) + \boldsymbol{\mathcal{W}}^T(\boldsymbol{x},\dot{\boldsymbol{x}})\right)\boldsymbol{\lambda} + \boldsymbol{\mathcal{Q}}^T_{\boldsymbol{\mathcal{S}}}(\boldsymbol{x})\,\boldsymbol{\mathcal{S}}\,(\boldsymbol{x} - \boldsymbol{x}_d(t)). \tag{4.241}$$

Using Eq. (4.227) to eliminate $\boldsymbol{\mathcal{P}}^T_{\boldsymbol{\mathcal{S}}}(\boldsymbol{x})\,\boldsymbol{\lambda}$ in Eqs. (4.240) and (4.241) results in

$$\begin{aligned}-\boldsymbol{\mathcal{P}}^T_{\boldsymbol{\mathcal{S}}}(\boldsymbol{x})\,\dot{\boldsymbol{\lambda}} = &\,\boldsymbol{\mathcal{P}}^T_{\boldsymbol{\mathcal{S}}}(\boldsymbol{x})\left(\nabla\boldsymbol{R}^T(\boldsymbol{x}) + \boldsymbol{\mathcal{W}}^T(\boldsymbol{x},\dot{\boldsymbol{x}})\right)\left(\boldsymbol{\mathcal{Q}}^T_{\boldsymbol{\mathcal{S}}}(\boldsymbol{x})\,\boldsymbol{\lambda} - \epsilon^2\boldsymbol{\Gamma}_{\boldsymbol{\mathcal{S}}}(\boldsymbol{x})\,(\dot{\boldsymbol{x}} - \boldsymbol{R}(\boldsymbol{x}))\right)\\ &+ \boldsymbol{\mathcal{P}}^T_{\boldsymbol{\mathcal{S}}}(\boldsymbol{x})\,\boldsymbol{\mathcal{S}}\,(\boldsymbol{x} - \boldsymbol{x}_d(t)),\end{aligned} \tag{4.242}$$

$$\begin{aligned}-\boldsymbol{\mathcal{Q}}^T_{\boldsymbol{\mathcal{S}}}(\boldsymbol{x})\,\dot{\boldsymbol{\lambda}} = &\,\boldsymbol{\mathcal{Q}}^T_{\boldsymbol{\mathcal{S}}}(\boldsymbol{x})\left(\nabla\boldsymbol{R}^T(\boldsymbol{x}) + \boldsymbol{\mathcal{W}}^T(\boldsymbol{x},\dot{\boldsymbol{x}})\right)\left(\boldsymbol{\mathcal{Q}}^T_{\boldsymbol{\mathcal{S}}}(\boldsymbol{x})\,\boldsymbol{\lambda} - \epsilon^2\boldsymbol{\Gamma}_{\boldsymbol{\mathcal{S}}}(\boldsymbol{x})\,(\dot{\boldsymbol{x}} - \boldsymbol{R}(\boldsymbol{x}))\right)\\ &+ \boldsymbol{\mathcal{Q}}^T_{\boldsymbol{\mathcal{S}}}(\boldsymbol{x})\,\boldsymbol{\mathcal{S}}\,(\boldsymbol{x} - \boldsymbol{x}_d(t)).\end{aligned} \tag{4.243}$$

Equations (4.230) and (4.242) are two independent expressions for $\boldsymbol{\mathcal{P}}^T_{\boldsymbol{\mathcal{S}}}(\boldsymbol{x})\,\dot{\boldsymbol{\lambda}}$. Combining them yields a second order differential equation independent of $\boldsymbol{\mathcal{P}}^T_{\boldsymbol{\mathcal{S}}}(\boldsymbol{x})\,\boldsymbol{\lambda}$,

$$\begin{aligned}\epsilon^2\boldsymbol{\Gamma}_{\boldsymbol{\mathcal{S}}}(\boldsymbol{x})\,\ddot{\boldsymbol{x}} = &\,\epsilon^2\boldsymbol{\Gamma}_{\boldsymbol{\mathcal{S}}}(\boldsymbol{x})\,\nabla\boldsymbol{R}(\boldsymbol{x})\,\dot{\boldsymbol{x}} - \epsilon^2\left(\dot{\boldsymbol{\Gamma}}_{\boldsymbol{\mathcal{S}}}(\boldsymbol{x}) - \dot{\boldsymbol{\mathcal{P}}}^T_{\boldsymbol{\mathcal{S}}}(\boldsymbol{x})\,\boldsymbol{\Gamma}_{\boldsymbol{\mathcal{S}}}(\boldsymbol{x})\right)(\dot{\boldsymbol{x}} - \boldsymbol{R}(\boldsymbol{x}))\\ &+ \boldsymbol{\mathcal{P}}^T_{\boldsymbol{\mathcal{S}}}(\boldsymbol{x})\left(\nabla\boldsymbol{R}^T(\boldsymbol{x}) + \boldsymbol{\mathcal{W}}^T(\boldsymbol{x},\dot{\boldsymbol{x}})\right)\left(\boldsymbol{\mathcal{Q}}^T_{\boldsymbol{\mathcal{S}}}(\boldsymbol{x})\,\boldsymbol{\lambda} - \epsilon^2\boldsymbol{\Gamma}_{\boldsymbol{\mathcal{S}}}(\boldsymbol{x})\,(\dot{\boldsymbol{x}} - \boldsymbol{R}(\boldsymbol{x}))\right)\\ &- \dot{\boldsymbol{\mathcal{P}}}^T_{\boldsymbol{\mathcal{S}}}(\boldsymbol{x})\,\boldsymbol{\mathcal{Q}}^T_{\boldsymbol{\mathcal{S}}}(\boldsymbol{x})\,\boldsymbol{\lambda} + \boldsymbol{\mathcal{P}}^T_{\boldsymbol{\mathcal{S}}}(\boldsymbol{x})\,\boldsymbol{\mathcal{S}}\,(\boldsymbol{x} - \boldsymbol{x}_d(t)).\end{aligned} \tag{4.244}$$

Equation (4.244) contains several time dependent matrices which can be simplified. From Eq. (4.215) follows for the time derivative of $\boldsymbol{\Gamma}_{\boldsymbol{\mathcal{S}}}(\boldsymbol{x})$

$$\dot{\boldsymbol{\Gamma}}_{\boldsymbol{\mathcal{S}}}(\boldsymbol{x}) = \dot{\boldsymbol{\mathcal{P}}}_{\boldsymbol{\mathcal{S}}}^{T}(\boldsymbol{x})\,\boldsymbol{\Gamma}_{\boldsymbol{\mathcal{S}}}(\boldsymbol{x}) + \boldsymbol{\mathcal{P}}_{\boldsymbol{\mathcal{S}}}^{T}(\boldsymbol{x})\,\dot{\boldsymbol{\Gamma}}_{\boldsymbol{\mathcal{S}}}(\boldsymbol{x}) \tag{4.245}$$

such that

$$\dot{\boldsymbol{\Gamma}}_{\boldsymbol{\mathcal{S}}}(\boldsymbol{x}) - \dot{\boldsymbol{\mathcal{P}}}_{\boldsymbol{\mathcal{S}}}^{T}(\boldsymbol{x})\,\boldsymbol{\Gamma}_{\boldsymbol{\mathcal{S}}}(\boldsymbol{x}) = \boldsymbol{\mathcal{P}}_{\boldsymbol{\mathcal{S}}}^{T}(\boldsymbol{x})\,\dot{\boldsymbol{\Gamma}}_{\boldsymbol{\mathcal{S}}}(\boldsymbol{x})\,. \tag{4.246}$$

Furthermore, from

$$\boldsymbol{\mathcal{P}}_{\boldsymbol{\mathcal{S}}}^{T}(\boldsymbol{x})\,\boldsymbol{\mathcal{Q}}_{\boldsymbol{\mathcal{S}}}^{T}(\boldsymbol{x}) = \boldsymbol{0} \tag{4.247}$$

follows

$$\dot{\boldsymbol{\mathcal{P}}}_{\boldsymbol{\mathcal{S}}}^{T}(\boldsymbol{x})\,\boldsymbol{\mathcal{Q}}_{\boldsymbol{\mathcal{S}}}^{T}(\boldsymbol{x}) = -\boldsymbol{\mathcal{P}}_{\boldsymbol{\mathcal{S}}}^{T}(\boldsymbol{x})\,\dot{\boldsymbol{\mathcal{Q}}}_{\boldsymbol{\mathcal{S}}}^{T}(\boldsymbol{x})\,, \tag{4.248}$$

and also

$$\dot{\boldsymbol{\mathcal{P}}}_{\boldsymbol{\mathcal{S}}}^{T}(\boldsymbol{x})\,\boldsymbol{\mathcal{Q}}_{\boldsymbol{\mathcal{S}}}^{T}(\boldsymbol{x})\,\boldsymbol{\mathcal{Q}}_{\boldsymbol{\mathcal{S}}}^{T}(\boldsymbol{x}) = -\boldsymbol{\mathcal{P}}_{\boldsymbol{\mathcal{S}}}^{T}(\boldsymbol{x})\,\dot{\boldsymbol{\mathcal{Q}}}_{\boldsymbol{\mathcal{S}}}^{T}(\boldsymbol{x})\,\boldsymbol{\mathcal{Q}}_{\boldsymbol{\mathcal{S}}}^{T}(\boldsymbol{x}) \tag{4.249}$$

due to idempotence of projectors. See also Section A.3 of the Appendix for more relations between time-dependent complementary projectors. Using Eqs. (4.246) and (4.248) in Eq. (4.244) yields

$$\begin{aligned}
\epsilon^2\boldsymbol{\Gamma}_{\boldsymbol{\mathcal{S}}}(\boldsymbol{x})\,\ddot{\boldsymbol{x}} &= \epsilon^2\boldsymbol{\mathcal{P}}_{\boldsymbol{\mathcal{S}}}^{T}(\boldsymbol{x})\left(\boldsymbol{\Gamma}_{\boldsymbol{\mathcal{S}}}(\boldsymbol{x})\,\nabla\boldsymbol{R}(\boldsymbol{x})\,\dot{\boldsymbol{x}} - \dot{\boldsymbol{\Gamma}}_{\boldsymbol{\mathcal{S}}}(\boldsymbol{x})\,(\dot{\boldsymbol{x}} - \boldsymbol{R}(\boldsymbol{x}))\right)\\
&\quad + \boldsymbol{\mathcal{P}}_{\boldsymbol{\mathcal{S}}}^{T}(\boldsymbol{x})\left(\nabla\boldsymbol{R}^{T}(\boldsymbol{x}) + \boldsymbol{\mathcal{W}}^{T}(\boldsymbol{x},\dot{\boldsymbol{x}})\right)\left(\boldsymbol{\mathcal{Q}}_{\boldsymbol{\mathcal{S}}}^{T}(\boldsymbol{x})\,\boldsymbol{\lambda} - \epsilon^2\boldsymbol{\Gamma}_{\boldsymbol{\mathcal{S}}}(\boldsymbol{x})\,(\dot{\boldsymbol{x}} - \boldsymbol{R}(\boldsymbol{x}))\right)\\
&\quad + \boldsymbol{\mathcal{P}}_{\boldsymbol{\mathcal{S}}}^{T}(\boldsymbol{x})\,\dot{\boldsymbol{\mathcal{Q}}}_{\boldsymbol{\mathcal{S}}}^{T}(\boldsymbol{x})\,\boldsymbol{\mathcal{Q}}_{\boldsymbol{\mathcal{S}}}^{T}(\boldsymbol{x})\,\boldsymbol{\lambda} + \boldsymbol{\mathcal{P}}_{\boldsymbol{\mathcal{S}}}^{T}(\boldsymbol{x})\,\boldsymbol{\mathcal{S}}\,(\boldsymbol{x} - \boldsymbol{x}_d(t))\,.
\end{aligned} \tag{4.250}$$

The form of Eq. (4.250) makes it obvious that it contains no component in the "direction" $\boldsymbol{\mathcal{Q}}_{\boldsymbol{\mathcal{S}}}^{T}(\boldsymbol{x})$. Equation (4.250) is a second order differential equation for $n-p$ independent state components $\boldsymbol{\mathcal{P}}_{\boldsymbol{\mathcal{S}}}(\boldsymbol{x})\,\boldsymbol{x}$. The $2\,(n-p)$ initial or terminal conditions necessary to solve Eq. (4.250) are given by Eqs. (4.219) and (4.220).

To summarize the derivation, the rearranged necessary optimality conditions are

$$\begin{aligned}
-\boldsymbol{\mathcal{Q}}_{\boldsymbol{\mathcal{S}}}^{T}(\boldsymbol{x})\,\dot{\boldsymbol{\lambda}} &= \boldsymbol{\mathcal{Q}}_{\boldsymbol{\mathcal{S}}}^{T}(\boldsymbol{x})\left(\nabla\boldsymbol{R}^{T}(\boldsymbol{x}) + \boldsymbol{\mathcal{W}}^{T}(\boldsymbol{x},\dot{\boldsymbol{x}})\right)\left(\boldsymbol{\mathcal{Q}}_{\boldsymbol{\mathcal{S}}}^{T}(\boldsymbol{x})\,\boldsymbol{\lambda} - \epsilon^2\boldsymbol{\Gamma}_{\boldsymbol{\mathcal{S}}}(\boldsymbol{x})\,(\dot{\boldsymbol{x}} - \boldsymbol{R}(\boldsymbol{x}))\right)\\
&\quad + \boldsymbol{\mathcal{Q}}_{\boldsymbol{\mathcal{S}}}^{T}(\boldsymbol{x})\,\boldsymbol{\mathcal{S}}\boldsymbol{\mathcal{Q}}_{\boldsymbol{\mathcal{S}}}(\boldsymbol{x})\,(\boldsymbol{x} - \boldsymbol{x}_d(t))\,,
\end{aligned} \tag{4.251}$$

$$\begin{aligned}
\epsilon^2\boldsymbol{\Gamma}_{\boldsymbol{\mathcal{S}}}(\boldsymbol{x})\,\ddot{\boldsymbol{x}} &= \epsilon^2\boldsymbol{\mathcal{P}}_{\boldsymbol{\mathcal{S}}}^{T}(\boldsymbol{x})\left(\boldsymbol{\Gamma}_{\boldsymbol{\mathcal{S}}}(\boldsymbol{x})\,\nabla\boldsymbol{R}(\boldsymbol{x})\,\dot{\boldsymbol{x}} - \dot{\boldsymbol{\Gamma}}_{\boldsymbol{\mathcal{S}}}(\boldsymbol{x})\,(\dot{\boldsymbol{x}} - \boldsymbol{R}(\boldsymbol{x}))\right)\\
&\quad + \boldsymbol{\mathcal{P}}_{\boldsymbol{\mathcal{S}}}^{T}(\boldsymbol{x})\left(\nabla\boldsymbol{R}^{T}(\boldsymbol{x}) + \boldsymbol{\mathcal{W}}^{T}(\boldsymbol{x},\dot{\boldsymbol{x}})\right)\left(\boldsymbol{\mathcal{Q}}_{\boldsymbol{\mathcal{S}}}^{T}(\boldsymbol{x})\,\boldsymbol{\lambda} - \epsilon^2\boldsymbol{\Gamma}_{\boldsymbol{\mathcal{S}}}(\boldsymbol{x})\,(\dot{\boldsymbol{x}} - \boldsymbol{R}(\boldsymbol{x}))\right)\\
&\quad + \boldsymbol{\mathcal{P}}_{\boldsymbol{\mathcal{S}}}^{T}(\boldsymbol{x})\,\dot{\boldsymbol{\mathcal{Q}}}_{\boldsymbol{\mathcal{S}}}^{T}(\boldsymbol{x})\,\boldsymbol{\mathcal{Q}}_{\boldsymbol{\mathcal{S}}}^{T}(\boldsymbol{x})\,\boldsymbol{\lambda} + \boldsymbol{\mathcal{P}}_{\boldsymbol{\mathcal{S}}}^{T}(\boldsymbol{x})\,\boldsymbol{\mathcal{S}}\boldsymbol{\mathcal{P}}_{\boldsymbol{\mathcal{S}}}(\boldsymbol{x})\,(\boldsymbol{x} - \boldsymbol{x}_d(t))\,,
\end{aligned} \tag{4.252}$$

$$\boldsymbol{\mathcal{Q}}_{\boldsymbol{\mathcal{S}}}(\boldsymbol{x})\,\dot{\boldsymbol{x}} = \boldsymbol{\mathcal{Q}}_{\boldsymbol{\mathcal{S}}}(\boldsymbol{x})\,\boldsymbol{R}(\boldsymbol{x})\,. \tag{4.253}$$

Equation (4.209) was used for the terms $\boldsymbol{\mathcal{Q}}_{\boldsymbol{\mathcal{S}}}^{T}\boldsymbol{\mathcal{S}}\boldsymbol{\mathcal{Q}}_{\boldsymbol{\mathcal{S}}}$ and $\boldsymbol{\mathcal{P}}_{\boldsymbol{\mathcal{S}}}^{T}\boldsymbol{\mathcal{S}}\boldsymbol{\mathcal{P}}_{\boldsymbol{\mathcal{S}}}$. We emphasize that these equations are just a rearrangement of the necessary optimality conditions Eqs. (4.198)-(4.200), and no approximation is involved. The small regularization parameter ϵ^2 multiplies the highest derivative $\ddot{\boldsymbol{x}}(t)$ in the system. This fact is exploited to perform a singular perturbation expansion.

4.4.2. Outer equations

The outer equations are obtained by expanding the rearranged necessary optimality conditions Eqs. (4.251)-(4.253) in ϵ. They are defined on the original time scale t. For the sake of clarity, the outer solutions are distinguished from the solutions $\boldsymbol{x}(t)$ and $\boldsymbol{\lambda}(t)$ by an index O,

$$\boldsymbol{x}_O(t) = \boldsymbol{x}(t), \qquad \boldsymbol{\lambda}_O(t) = \boldsymbol{\lambda}(t). \tag{4.254}$$

To shorten the notation, the time argument is suppressed in this subsection. Setting $\epsilon = 0$ in Eqs. (4.251)-(4.253) yields a system of algebraic and first order differential equations,

$$\begin{aligned}
-\boldsymbol{\mathcal{Q}}_{\boldsymbol{\mathcal{S}}}^T(\boldsymbol{x}_O)\,\dot{\boldsymbol{\lambda}}_O &= \boldsymbol{\mathcal{Q}}_{\boldsymbol{\mathcal{S}}}^T(\boldsymbol{x}_O)\left(\nabla \boldsymbol{R}^T(\boldsymbol{x}_O) + \boldsymbol{\mathcal{W}}^T(\boldsymbol{x}_O, \dot{\boldsymbol{x}}_O)\right)\boldsymbol{\mathcal{Q}}_{\boldsymbol{\mathcal{S}}}^T(\boldsymbol{x}_O)\,\boldsymbol{\lambda}_O \\
&\quad + \boldsymbol{\mathcal{Q}}_{\boldsymbol{\mathcal{S}}}^T(\boldsymbol{x}_O)\,\boldsymbol{\mathcal{S}}\boldsymbol{\mathcal{Q}}_{\boldsymbol{\mathcal{S}}}(\boldsymbol{x}_O)\,(\boldsymbol{x}_O - \boldsymbol{x}_d(t)), \qquad (4.255)\\
\boldsymbol{\mathcal{P}}_{\boldsymbol{\mathcal{S}}}^T(\boldsymbol{x}_O)\,\boldsymbol{\mathcal{S}}\boldsymbol{\mathcal{P}}_{\boldsymbol{\mathcal{S}}}(\boldsymbol{x}_O)\,\boldsymbol{x}_O &= \boldsymbol{\mathcal{P}}_{\boldsymbol{\mathcal{S}}}^T(\boldsymbol{x}_O)\,\boldsymbol{\mathcal{S}}\boldsymbol{\mathcal{P}}_{\boldsymbol{\mathcal{S}}}(\boldsymbol{x}_O)\,\boldsymbol{x}_d(t) \\
&\quad - \boldsymbol{\mathcal{P}}_{\boldsymbol{\mathcal{S}}}^T(\boldsymbol{x}_O)\left(\nabla \boldsymbol{R}^T(\boldsymbol{x}_O) + \boldsymbol{\mathcal{W}}^T(\boldsymbol{x}_O, \dot{\boldsymbol{x}}_O)\right)\boldsymbol{\mathcal{Q}}_{\boldsymbol{\mathcal{S}}}^T(\boldsymbol{x}_O)\,\boldsymbol{\lambda}_O \\
&\quad - \boldsymbol{\mathcal{P}}_{\boldsymbol{\mathcal{S}}}^T(\boldsymbol{x}_O)\,\dot{\boldsymbol{\mathcal{Q}}}_{\boldsymbol{\mathcal{S}}}^T(\boldsymbol{x}_O)\,\boldsymbol{\mathcal{Q}}_{\boldsymbol{\mathcal{S}}}^T(\boldsymbol{x}_O)\,\boldsymbol{\lambda}_O, \qquad (4.256)\\
\boldsymbol{\mathcal{Q}}_{\boldsymbol{\mathcal{S}}}(\boldsymbol{x}_O)\,\dot{\boldsymbol{x}}_O &= \boldsymbol{\mathcal{Q}}_{\boldsymbol{\mathcal{S}}}(\boldsymbol{x}_O)\,\boldsymbol{R}(\boldsymbol{x}_O). \qquad (4.257)
\end{aligned}$$

Equation (4.256) is used to obtain an expression for $\boldsymbol{\mathcal{P}}_{\boldsymbol{\mathcal{S}}}(\boldsymbol{x}_O)\,\boldsymbol{x}_O$. Multiplying Eq. (4.256) from the left by $\boldsymbol{\Omega}_{\boldsymbol{\mathcal{S}}}(\boldsymbol{x}_O)$ and exploiting Eqs. (4.210) and (4.211) yields

$$\begin{aligned}
\boldsymbol{\mathcal{P}}_{\boldsymbol{\mathcal{S}}}(\boldsymbol{x}_O)\,\boldsymbol{x}_O &= -\boldsymbol{\Omega}_{\boldsymbol{\mathcal{S}}}(\boldsymbol{x}_O)\left(\dot{\boldsymbol{\mathcal{Q}}}_{\boldsymbol{\mathcal{S}}}^T(\boldsymbol{x}_O) + \nabla \boldsymbol{R}^T(\boldsymbol{x}_O) + \boldsymbol{\mathcal{W}}^T(\boldsymbol{x}_O, \dot{\boldsymbol{x}}_O)\right)\boldsymbol{\mathcal{Q}}_{\boldsymbol{\mathcal{S}}}^T(\boldsymbol{x}_O)\,\boldsymbol{\lambda}_O \\
&\quad + \boldsymbol{\mathcal{P}}_{\boldsymbol{\mathcal{S}}}(\boldsymbol{x}_O)\,\boldsymbol{x}_d(t). \qquad (4.258)
\end{aligned}$$

Equation (4.258) is not a closed form expression for $\boldsymbol{\mathcal{P}}_{\boldsymbol{\mathcal{S}}}(\boldsymbol{x}_O)\,\boldsymbol{x}_O$ as long as $\boldsymbol{\Omega}_{\boldsymbol{\mathcal{S}}}(\boldsymbol{x}_O)$, $\boldsymbol{\mathcal{W}}^T(\boldsymbol{x}_O, \dot{\boldsymbol{x}}_O)$, and $\nabla \boldsymbol{R}^T(\boldsymbol{x}_O)$ depend on $\boldsymbol{\mathcal{P}}_{\boldsymbol{\mathcal{S}}}(\boldsymbol{x}_O)\,\boldsymbol{x}_O$ as well. Nevertheless, Eq. (4.258) can be used to eliminate $\boldsymbol{\mathcal{P}}_{\boldsymbol{\mathcal{S}}}(\boldsymbol{x}_O)\,\boldsymbol{x}_O$ in Eqs. (4.255) and (4.257). The derivation of the explicit expression for $\boldsymbol{\mathcal{P}}_{\boldsymbol{\mathcal{S}}}(\boldsymbol{x}_O)\,\boldsymbol{x}_O$ in Eq. (4.258) relies on the usage of the projectors $\boldsymbol{\mathcal{P}}_{\boldsymbol{\mathcal{S}}}$ and $\boldsymbol{\mathcal{Q}}_{\boldsymbol{\mathcal{S}}}$. It is impossible to derive an analogous relation using the projectors $\boldsymbol{\mathcal{P}}$ and $\boldsymbol{\mathcal{Q}}$ defined in Chapter 2 instead of $\boldsymbol{\mathcal{P}}_{\boldsymbol{\mathcal{S}}}$ and $\boldsymbol{\mathcal{Q}}_{\boldsymbol{\mathcal{S}}}$. In particular, $\boldsymbol{\mathcal{P}}$ and $\boldsymbol{\mathcal{Q}}$ cannot satisfy relations analogous to Eqs. (4.210), (4.211), and (4.209). This motivates the usage of the projectors $\boldsymbol{\mathcal{P}}_{\boldsymbol{\mathcal{S}}}$ and $\boldsymbol{\mathcal{Q}}_{\boldsymbol{\mathcal{S}}}$ retrospectively.

4.4.3. Inner equations - left side

Boundary layers are expected at the left and right hand side of the time domain. An initial boundary layer at the left end is resolved by the time scale τ_L defined as

$$\tau_L = (t - t_0)/\epsilon^{\alpha}. \tag{4.259}$$

The exponent α has to be determined by dominant balance of the leading order terms as $\epsilon \to 0$ (Bender and Orszag, 2010; Johnson, 2004). The left inner solutions are denoted by capital letters with index L,

$$\boldsymbol{X}_L(\tau_L) = \boldsymbol{X}_L((t-t_0)/\epsilon^\alpha) = \boldsymbol{x}(t), \quad \boldsymbol{\Lambda}_L(\tau_L) = \boldsymbol{\Lambda}_L((t-t_0)/\epsilon^\alpha) = \boldsymbol{\lambda}(t). \tag{4.260}$$

Expressed in terms of the inner solutions, the time derivatives become

$$\dot{\boldsymbol{x}}(t) = \frac{d}{dt}\boldsymbol{X}_L((t-t_0)/\epsilon^\alpha) = \epsilon^{-\alpha}\boldsymbol{X}'_L(\tau_L), \qquad \dot{\boldsymbol{\lambda}}(t) = \epsilon^{-\alpha}\boldsymbol{\Lambda}'_L(\tau_L), \tag{4.261}$$

$$\ddot{\boldsymbol{x}}(t) = \frac{d^2}{dt^2}\boldsymbol{X}_L((t-t_0)/\epsilon^\alpha) = \epsilon^{-2\alpha}\boldsymbol{X}''_L(\tau_L). \tag{4.262}$$

The prime $\boldsymbol{X}'_L(\tau_L)$ denotes the derivative of $\boldsymbol{X}_L$ with respect to its argument τ_L. The time derivatives of $\boldsymbol{\mathcal{Q}}_{\boldsymbol{\mathcal{S}}}(\boldsymbol{x}(t))$ and $\boldsymbol{\Gamma}_{\boldsymbol{\mathcal{S}}}(\boldsymbol{x}(t))$ transform as

$$\dot{\boldsymbol{\mathcal{Q}}}^T_{\boldsymbol{\mathcal{S}}}(\boldsymbol{x}(t)) = \epsilon^{-\alpha}\nabla\boldsymbol{\mathcal{Q}}^T_{\boldsymbol{\mathcal{S}}}(\boldsymbol{X}_L(\tau_L))\boldsymbol{X}'_L(\tau_L) = \epsilon^{-\alpha}\boldsymbol{\mathcal{Q}}^{T\prime}_{\boldsymbol{\mathcal{S}}}(\boldsymbol{X}_L(\tau_L)), \tag{4.263}$$

$$\dot{\boldsymbol{\Gamma}}_{\boldsymbol{\mathcal{S}}}(\boldsymbol{x}(t)) = \epsilon^{-\alpha}\nabla\boldsymbol{\Gamma}_{\boldsymbol{\mathcal{S}}}(\boldsymbol{X}_L(\tau_L))\boldsymbol{X}'_L(\tau_L) = \epsilon^{-\alpha}\boldsymbol{\Gamma}'_{\boldsymbol{\mathcal{S}}}(\boldsymbol{X}_L(\tau_L)). \tag{4.264}$$

To shorten the notation, the prime on the matrix is defined as

$$\boldsymbol{\Gamma}'_{\boldsymbol{\mathcal{S}}}(\boldsymbol{X}_L(\tau_L)) = \nabla\boldsymbol{\Gamma}_{\boldsymbol{\mathcal{S}}}(\boldsymbol{X}_L(\tau_L))\boldsymbol{X}'_L(\tau_L). \tag{4.265}$$

The matrix $\boldsymbol{\mathcal{W}}(\boldsymbol{x}(t), \dot{\boldsymbol{x}}(t))$ transforms as

$$\begin{aligned}\boldsymbol{\mathcal{W}}(\boldsymbol{x}(t), \dot{\boldsymbol{x}}(t)) &= \nabla\boldsymbol{\mathcal{B}}(\boldsymbol{x}(t))\boldsymbol{\mathcal{B}}^g_{\boldsymbol{\mathcal{S}}}(\boldsymbol{x}(t))\dot{\boldsymbol{x}}(t) - \nabla\boldsymbol{\mathcal{B}}(\boldsymbol{x}(t))\boldsymbol{\mathcal{B}}^g_{\boldsymbol{\mathcal{S}}}(\boldsymbol{x}(t))\boldsymbol{R}(\boldsymbol{x}(t)) \\ &= \epsilon^{-\alpha}\boldsymbol{\mathcal{V}}(\boldsymbol{X}_L(\tau_L), \boldsymbol{X}'_L(\tau_L)) + \boldsymbol{\mathcal{U}}(\boldsymbol{X}_L(\tau_L)),\end{aligned} \tag{4.266}$$

with $n \times n$ matrices $\boldsymbol{\mathcal{U}}$ and $\boldsymbol{\mathcal{V}}$ defined by

$$\boldsymbol{\mathcal{U}}(\boldsymbol{x}) = -\nabla\boldsymbol{\mathcal{B}}(\boldsymbol{x})\boldsymbol{\mathcal{B}}^g_{\boldsymbol{\mathcal{S}}}(\boldsymbol{x})\boldsymbol{R}(\boldsymbol{x}), \tag{4.267}$$

$$\boldsymbol{\mathcal{V}}(\boldsymbol{x}, \boldsymbol{y}) = \nabla\boldsymbol{\mathcal{B}}(\boldsymbol{x})\boldsymbol{\mathcal{B}}^g_{\boldsymbol{\mathcal{S}}}(\boldsymbol{x})\boldsymbol{y}. \tag{4.268}$$

The entries of $\boldsymbol{\mathcal{U}}$ and $\boldsymbol{\mathcal{V}}$ are

$$\mathcal{U}_{ij}(\boldsymbol{x}) = \sum_{k=1}^{n}\sum_{l=1}^{p}\frac{\partial}{\partial x_j}\mathcal{B}_{il}(\boldsymbol{x})\mathcal{B}^g_{\boldsymbol{\mathcal{S}},lk}(\boldsymbol{x})R_k(\boldsymbol{x}), \tag{4.269}$$

$$\mathcal{V}_{ij}(\boldsymbol{x}, \boldsymbol{y}) = \sum_{k=1}^{n}\sum_{l=1}^{p}\frac{\partial}{\partial x_j}\mathcal{B}_{il}(\boldsymbol{x})\mathcal{B}^g_{\boldsymbol{\mathcal{S}},lk}(\boldsymbol{x})y_k. \tag{4.270}$$

From the initial conditions Eq. (4.201) follow the initial conditions for $\boldsymbol{X}_L(\tau_L)$ as

$$\boldsymbol{X}_L(0) = \boldsymbol{x}_0. \tag{4.271}$$

Transforming the necessary optimality conditions Eqs. (4.251)-(4.253) yields

$$\begin{aligned}
-\epsilon^{-\alpha}\boldsymbol{\mathcal{Q}}_{\boldsymbol{\mathcal{S}}}^{T}\left(\boldsymbol{X}_{L}\right)\boldsymbol{\Lambda}_{L}' = &-\epsilon^{2-2\alpha}\boldsymbol{\mathcal{Q}}_{\boldsymbol{\mathcal{S}}}^{T}\left(\boldsymbol{X}_{L}\right)\boldsymbol{\mathcal{V}}^{T}\left(\boldsymbol{X}_{L},\boldsymbol{X}_{L}'\right)\boldsymbol{\Gamma}_{\boldsymbol{\mathcal{S}}}\left(\boldsymbol{X}_{L}\right)\boldsymbol{X}_{L}' \\
&+\epsilon^{2-\alpha}\boldsymbol{\mathcal{Q}}_{\boldsymbol{\mathcal{S}}}^{T}\left(\boldsymbol{X}_{L}\right)\boldsymbol{\mathcal{V}}^{T}\left(\boldsymbol{X}_{L},\boldsymbol{X}_{L}'\right)\boldsymbol{\Gamma}_{\boldsymbol{\mathcal{S}}}\left(\boldsymbol{X}_{L}\right)\boldsymbol{R}\left(\boldsymbol{X}_{L}\right) \\
&-\epsilon^{2-\alpha}\boldsymbol{\mathcal{Q}}_{\boldsymbol{\mathcal{S}}}^{T}\left(\boldsymbol{X}_{L}\right)\left(\nabla\boldsymbol{R}^{T}\left(\boldsymbol{X}_{L}\right)+\boldsymbol{\mathcal{U}}^{T}\left(\boldsymbol{X}_{L}\right)\right)\boldsymbol{\Gamma}_{\boldsymbol{\mathcal{S}}}\left(\boldsymbol{X}_{L}\right)\boldsymbol{X}_{L}' \\
&+\epsilon^{-\alpha}\boldsymbol{\mathcal{Q}}_{\boldsymbol{\mathcal{S}}}^{T}\left(\boldsymbol{X}_{L}\right)\boldsymbol{\mathcal{V}}^{T}\left(\boldsymbol{X}_{L},\boldsymbol{X}_{L}'\right)\boldsymbol{\mathcal{Q}}_{\boldsymbol{\mathcal{S}}}^{T}\left(\boldsymbol{X}_{L}\right)\boldsymbol{\Lambda}_{L} \\
&+\epsilon^{2}\boldsymbol{\mathcal{Q}}_{\boldsymbol{\mathcal{S}}}^{T}\left(\boldsymbol{X}_{L}\right)\left(\nabla\boldsymbol{R}^{T}\left(\boldsymbol{X}_{L}\right)+\boldsymbol{\mathcal{U}}^{T}\left(\boldsymbol{X}_{L}\right)\right)\boldsymbol{\Gamma}_{\boldsymbol{\mathcal{S}}}\left(\boldsymbol{X}_{L}\right)\boldsymbol{R}\left(\boldsymbol{X}_{L}\right) \\
&+\boldsymbol{\mathcal{Q}}_{\boldsymbol{\mathcal{S}}}^{T}\left(\boldsymbol{X}_{L}\right)\left(\nabla\boldsymbol{R}^{T}\left(\boldsymbol{X}_{L}\right)+\boldsymbol{\mathcal{U}}^{T}\left(\boldsymbol{X}_{L}\right)\right)\boldsymbol{\mathcal{Q}}_{\boldsymbol{\mathcal{S}}}^{T}\left(\boldsymbol{X}_{L}\right)\boldsymbol{\Lambda}_{L} \\
&+\boldsymbol{\mathcal{Q}}_{\boldsymbol{\mathcal{S}}}^{T}\left(\boldsymbol{X}_{L}\right)\boldsymbol{\mathcal{S}}\boldsymbol{\mathcal{Q}}_{\boldsymbol{\mathcal{S}}}\left(\boldsymbol{X}_{L}\right)\left(\boldsymbol{X}_{L}-\boldsymbol{x}_{d}\left(t_{0}+\epsilon^{\alpha}\tau_{L}\right)\right), \qquad (4.272)
\end{aligned}$$

$$\begin{aligned}
\epsilon^{2-2\alpha}\boldsymbol{\Gamma}_{\boldsymbol{\mathcal{S}}}\left(\boldsymbol{X}_{L}\right)\boldsymbol{X}_{L}'' = &-\epsilon^{2-2\alpha}\boldsymbol{\mathcal{P}}_{\boldsymbol{\mathcal{S}}}^{T}\left(\boldsymbol{X}_{L}\right)\left(\boldsymbol{\mathcal{V}}^{T}\left(\boldsymbol{X}_{L},\boldsymbol{X}_{L}'\right)\boldsymbol{\Gamma}_{\boldsymbol{\mathcal{S}}}\left(\boldsymbol{X}_{L}\right)+\boldsymbol{\Gamma}_{\boldsymbol{\mathcal{S}}}'\left(\boldsymbol{X}_{L}\right)\right)\boldsymbol{X}_{L}' \\
&-\epsilon^{2-\alpha}\boldsymbol{\mathcal{P}}_{\boldsymbol{\mathcal{S}}}^{T}\left(\boldsymbol{X}_{L}\right)\boldsymbol{\Gamma}_{\boldsymbol{\mathcal{S}}}\left(\boldsymbol{X}_{L}\right)\nabla\boldsymbol{R}\left(\boldsymbol{X}_{L}\right)\boldsymbol{X}_{L}'\left(\tau_{L}\right) \\
&+\epsilon^{2-\alpha}\boldsymbol{\mathcal{P}}_{\boldsymbol{\mathcal{S}}}^{T}\left(\boldsymbol{X}_{L}\right)\boldsymbol{\Gamma}_{\boldsymbol{\mathcal{S}}}'\left(\boldsymbol{X}_{L}\right)\boldsymbol{R}\left(\boldsymbol{X}_{L}\right) \\
&-\epsilon^{2-\alpha}\boldsymbol{\mathcal{P}}_{\boldsymbol{\mathcal{S}}}^{T}\left(\boldsymbol{X}_{L}\right)\left(\nabla\boldsymbol{R}^{T}\left(\boldsymbol{X}_{L}\right)+\boldsymbol{\mathcal{U}}^{T}\left(\boldsymbol{X}_{L}\right)\right)\boldsymbol{\Gamma}_{\boldsymbol{\mathcal{S}}}\left(\boldsymbol{X}_{L}\right)\boldsymbol{X}_{L}' \\
&+\epsilon^{2-\alpha}\boldsymbol{\mathcal{P}}_{\boldsymbol{\mathcal{S}}}^{T}\left(\boldsymbol{X}_{L}\right)\boldsymbol{\mathcal{V}}^{T}\left(\boldsymbol{X}_{L},\boldsymbol{X}_{L}'\right)\boldsymbol{\Gamma}_{\boldsymbol{\mathcal{S}}}\left(\boldsymbol{X}_{L}\right)\boldsymbol{R}\left(\boldsymbol{X}_{L}\right) \\
&+\epsilon^{-\alpha}\boldsymbol{\mathcal{P}}_{\boldsymbol{\mathcal{S}}}^{T}\left(\boldsymbol{X}_{L}\right)\left(\boldsymbol{\mathcal{V}}^{T}\left(\boldsymbol{X}_{L},\boldsymbol{X}_{L}'\right)+\boldsymbol{\mathcal{Q}}_{\boldsymbol{\mathcal{S}}}^{T\prime}\left(\boldsymbol{X}_{L}\right)\right)\boldsymbol{\mathcal{Q}}_{\boldsymbol{\mathcal{S}}}^{T}\left(\boldsymbol{X}_{L}\right)\boldsymbol{\Lambda}_{L} \\
&+\epsilon^{2}\boldsymbol{\mathcal{P}}_{\boldsymbol{\mathcal{S}}}^{T}\left(\boldsymbol{X}_{L}\right)\left(\nabla\boldsymbol{R}^{T}\left(\boldsymbol{X}_{L}\right)+\boldsymbol{\mathcal{U}}^{T}\left(\boldsymbol{X}_{L}\right)\right)\boldsymbol{\Gamma}_{\boldsymbol{\mathcal{S}}}\left(\boldsymbol{X}_{L}\right)\boldsymbol{R}\left(\boldsymbol{X}_{L}\right) \\
&+\boldsymbol{\mathcal{P}}_{\boldsymbol{\mathcal{S}}}^{T}\left(\boldsymbol{X}_{L}\right)\left(\nabla\boldsymbol{R}^{T}\left(\boldsymbol{X}_{L}\right)+\boldsymbol{\mathcal{U}}^{T}\left(\boldsymbol{X}_{L}\right)\right)\boldsymbol{\mathcal{Q}}_{\boldsymbol{\mathcal{S}}}^{T}\left(\boldsymbol{X}_{L}\right)\boldsymbol{\Lambda}_{L} \\
&+\boldsymbol{\mathcal{P}}_{\boldsymbol{\mathcal{S}}}^{T}\left(\boldsymbol{X}_{L}\right)\boldsymbol{\mathcal{S}}\boldsymbol{\mathcal{P}}_{\boldsymbol{\mathcal{S}}}\left(\boldsymbol{X}_{L}\right)\left(\boldsymbol{X}_{L}-\boldsymbol{x}_{d}\left(t_{0}+\epsilon^{\alpha}\tau_{L}\right)\right), \qquad (4.273)
\end{aligned}$$

$$\epsilon^{-\alpha}\boldsymbol{\mathcal{Q}}_{\boldsymbol{\mathcal{S}}}\left(\boldsymbol{X}_{L}\right)\boldsymbol{X}_{L}' = \boldsymbol{\mathcal{Q}}_{\boldsymbol{\mathcal{S}}}\left(\boldsymbol{X}_{L}\right)\boldsymbol{R}\left(\boldsymbol{X}_{L}\right). \qquad (4.274)$$

A dominant balance argument is applied to determine the possible values of the exponent α. Collecting the exponents of ϵ yields a list

$$2-2\alpha, \quad 2-\alpha, \quad -\alpha, \quad 2, \quad 0. \qquad (4.275)$$

A dominant balance occurs if at least one pair of equal exponents exists (Bender and Orszag, 2010). Pairs of equal exponents appear as intersections of straight lines in a plot of the exponent values (4.275) over α, see Fig. 4.13.

The value $\alpha = 0$ leads to the outer equations and is discarded. The relevant values of α as given by dominant balance are $\alpha = 2$, $\alpha = 1$, and $\alpha = -2$. A case by case analysis is performed in the following.

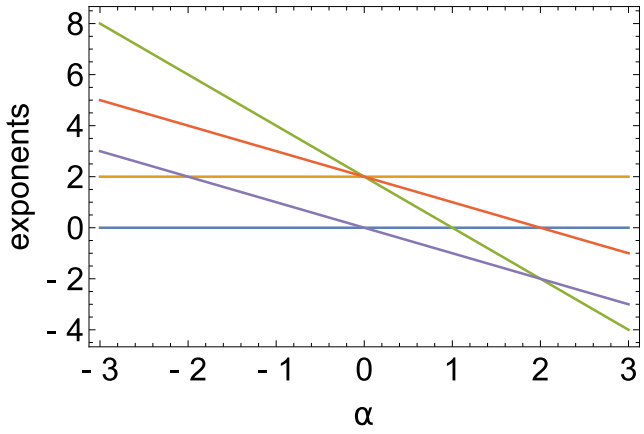

Figure 4.13.: The values of exponents of ϵ are plotted over α. An intersection indicates a dominant balance and determines a possible value for α.

4.4.3.1. Case 1 with $\alpha = 2$

The leading order equations are $2\,(n-p)$ first order differential equations and p second order differential equations,

$$\boldsymbol{\mathcal{Q}}_{\boldsymbol{\mathcal{S}}}^{T}\left(\boldsymbol{X}_{L}\right)\boldsymbol{\Lambda}_{L}^{\prime}=\boldsymbol{\mathcal{Q}}_{\boldsymbol{\mathcal{S}}}^{T}\left(\boldsymbol{X}_{L}\right)\boldsymbol{\mathcal{V}}^{T}\left(\boldsymbol{X}_{L},\boldsymbol{X}_{L}^{\prime}\right)\left(\boldsymbol{\Gamma}_{\boldsymbol{\mathcal{S}}}\left(\boldsymbol{X}_{L}\right)\boldsymbol{X}_{L}^{\prime}-\boldsymbol{\mathcal{Q}}_{\boldsymbol{\mathcal{S}}}^{T}\left(\boldsymbol{X}_{L}\right)\boldsymbol{\Lambda}_{L}\right) \tag{4.276}$$

$$\begin{aligned}\frac{\partial}{\partial\tau_{L}}\left(\boldsymbol{\Gamma}_{\boldsymbol{\mathcal{S}}}\left(\boldsymbol{X}_{L}\right)\boldsymbol{X}_{L}^{\prime}\right)&=\boldsymbol{\mathcal{P}}_{\boldsymbol{\mathcal{S}}}^{T}\left(\boldsymbol{X}_{L}\right)\left(\boldsymbol{\mathcal{V}}^{T}\left(\boldsymbol{X}_{L},\boldsymbol{X}_{L}^{\prime}\right)+\boldsymbol{\mathcal{Q}}_{\boldsymbol{\mathcal{S}}}^{T\prime}\left(\boldsymbol{X}_{L}\right)\right)\boldsymbol{\mathcal{Q}}_{\boldsymbol{\mathcal{S}}}^{T}\left(\boldsymbol{X}_{L}\right)\boldsymbol{\Lambda}_{L}\\&\quad-\boldsymbol{\mathcal{P}}_{\boldsymbol{\mathcal{S}}}^{T}\left(\boldsymbol{X}_{L}\right)\boldsymbol{\mathcal{V}}^{T}\left(\boldsymbol{X}_{L},\boldsymbol{X}_{L}^{\prime}\right)\boldsymbol{\Gamma}_{\boldsymbol{\mathcal{S}}}\left(\boldsymbol{X}_{L}\right)\boldsymbol{X}_{L}^{\prime},\end{aligned} \tag{4.277}$$

$$\boldsymbol{\mathcal{Q}}_{\boldsymbol{\mathcal{S}}}\left(\boldsymbol{X}_{L}\right)\boldsymbol{X}_{L}^{\prime}=\boldsymbol{0}. \tag{4.278}$$

Equation (4.277) was simplified by using the identity

$$\frac{\partial}{\partial\tau_{L}}\left(\boldsymbol{\Gamma}_{\boldsymbol{\mathcal{S}}}\left(\boldsymbol{X}_{L}\right)\boldsymbol{X}_{L}^{\prime}\right)=\boldsymbol{\Gamma}_{\boldsymbol{\mathcal{S}}}^{\prime}\left(\boldsymbol{X}_{L}\right)\boldsymbol{X}_{L}^{\prime}+\boldsymbol{\mathcal{P}}_{\boldsymbol{\mathcal{S}}}^{T}\left(\boldsymbol{X}_{L}\right)\boldsymbol{\Gamma}_{\boldsymbol{\mathcal{S}}}\left(\boldsymbol{X}_{L}\right)\boldsymbol{X}_{L}^{\prime\prime} \tag{4.279}$$

together with Eq. (4.215).

4.4.3.2. Case 2.1 with $\alpha = 1$

As long as

$$\boldsymbol{\mathcal{P}}_{\boldsymbol{\mathcal{S}}}^{T}\left(\boldsymbol{X}_{L}\right)\left(\boldsymbol{\mathcal{V}}^{T}\left(\boldsymbol{X}_{L},\boldsymbol{X}_{L}^{\prime}\right)+\boldsymbol{\mathcal{Q}}_{\boldsymbol{\mathcal{S}}}^{T\prime}\left(\boldsymbol{X}_{L}\right)\right)\boldsymbol{\mathcal{Q}}_{\boldsymbol{\mathcal{S}}}^{T}\left(\boldsymbol{X}_{L}\right)\boldsymbol{\Lambda}_{L}\neq\boldsymbol{0}, \tag{4.280}$$

the leading order equations are $2\,(n-p)+p$ first order differential equations,

$$\boldsymbol{\mathcal{Q}}_{\boldsymbol{\mathcal{S}}}^{T}\left(\boldsymbol{X}_{L}\right)\boldsymbol{\Lambda}_{L}^{\prime}=-\boldsymbol{\mathcal{Q}}_{\boldsymbol{\mathcal{S}}}^{T}\left(\boldsymbol{X}_{L}\right)\boldsymbol{\mathcal{V}}^{T}\left(\boldsymbol{X}_{L},\boldsymbol{X}_{L}^{\prime}\right)\boldsymbol{\mathcal{Q}}_{\boldsymbol{\mathcal{S}}}^{T}\left(\boldsymbol{X}_{L}\right)\boldsymbol{\Lambda}_{L}, \tag{4.281}$$

$$\boldsymbol{0}=\boldsymbol{\mathcal{P}}_{\boldsymbol{\mathcal{S}}}^{T}\left(\boldsymbol{X}_{L}\right)\left(\boldsymbol{\mathcal{V}}^{T}\left(\boldsymbol{X}_{L},\boldsymbol{X}_{L}^{\prime}\right)+\boldsymbol{\mathcal{Q}}_{\boldsymbol{\mathcal{S}}}^{T\prime}\left(\boldsymbol{X}_{L}\right)\right)\boldsymbol{\mathcal{Q}}_{\boldsymbol{\mathcal{S}}}^{T}\left(\boldsymbol{X}_{L}\right)\boldsymbol{\Lambda}_{L}, \tag{4.282}$$

$$\boldsymbol{\mathcal{Q}}_{\boldsymbol{\mathcal{S}}}\left(\boldsymbol{X}_{L}\right)\boldsymbol{X}_{L}^{\prime}=\boldsymbol{0}. \tag{4.283}$$

4.4.3.3. Case 2.2 with $\alpha = 1$

The leading order equations for $\alpha = 1$ change if

$$\boldsymbol{\mathcal{P}}_{\boldsymbol{\mathcal{S}}}^{T}\left(\boldsymbol{X}_{L}\right)\left(\boldsymbol{\mathcal{V}}^{T}\left(\boldsymbol{X}_{L},\boldsymbol{X}_{L}^{\prime}\right)+\boldsymbol{\mathcal{Q}}_{\boldsymbol{\mathcal{S}}}^{T\prime}\left(\boldsymbol{X}_{L}\right)\right)\boldsymbol{\mathcal{Q}}_{\boldsymbol{\mathcal{S}}}^{T}\left(\boldsymbol{X}_{L}\right)\boldsymbol{\Lambda}_{L}=\mathbf{0}. \tag{4.284}$$

In this case, the leading order equations are $2\left(n-p\right)$ first order differential equations and p second order differential equations,

$$\boldsymbol{\mathcal{Q}}_{\boldsymbol{\mathcal{S}}}^{T}\left(\boldsymbol{X}_{L}\right)\boldsymbol{\Lambda}_{L}^{\prime}=-\boldsymbol{\mathcal{Q}}_{\boldsymbol{\mathcal{S}}}^{T}\left(\boldsymbol{X}_{L}\right)\boldsymbol{\mathcal{V}}^{T}\left(\boldsymbol{X}_{L},\boldsymbol{X}_{L}^{\prime}\right)\boldsymbol{\mathcal{Q}}_{\boldsymbol{\mathcal{S}}}^{T}\left(\boldsymbol{X}_{L}\right)\boldsymbol{\Lambda}_{L}, \tag{4.285}$$

$$\begin{aligned}\frac{\partial}{\partial\tau_{L}}\left(\boldsymbol{\Gamma}_{\boldsymbol{\mathcal{S}}}\left(\boldsymbol{X}_{L}\right)\boldsymbol{X}_{L}^{\prime}\right)=&-\boldsymbol{\mathcal{P}}_{\boldsymbol{\mathcal{S}}}^{T}\left(\boldsymbol{X}_{L}\right)\boldsymbol{\mathcal{V}}^{T}\left(\boldsymbol{X}_{L},\boldsymbol{X}_{L}^{\prime}\right)\boldsymbol{\Gamma}_{\boldsymbol{\mathcal{S}}}\left(\boldsymbol{X}_{L}\right)\boldsymbol{X}_{L}^{\prime}\\&+\boldsymbol{\mathcal{P}}_{\boldsymbol{\mathcal{S}}}^{T}\left(\boldsymbol{X}_{L}\right)\left(\nabla\boldsymbol{R}^{T}\left(\boldsymbol{X}_{L}\right)+\boldsymbol{\mathcal{U}}^{T}\left(\boldsymbol{X}_{L}\right)\right)\boldsymbol{\mathcal{Q}}_{\boldsymbol{\mathcal{S}}}^{T}\left(\boldsymbol{X}_{L}\right)\boldsymbol{\Lambda}_{L}\\&+\boldsymbol{\mathcal{P}}_{\boldsymbol{\mathcal{S}}}^{T}\left(\boldsymbol{X}_{L}\right)\boldsymbol{\mathcal{S}}\boldsymbol{\mathcal{P}}_{\boldsymbol{\mathcal{S}}}\left(\boldsymbol{X}_{L}\right)\left(\boldsymbol{X}_{L}-\boldsymbol{x}_{d}\left(t_{0}\right)\right),\end{aligned} \tag{4.286}$$

$$\boldsymbol{\mathcal{Q}}_{\boldsymbol{\mathcal{S}}}\left(\boldsymbol{X}_{L}\right)\boldsymbol{X}_{L}^{\prime}=\mathbf{0}. \tag{4.287}$$

This case corresponds to the left inner equations of the two-dimensional dynamical system from Section 4.1. Equations (4.285)-(4.287) reduce to the left inner equations derived in Section 4.1.5.1.

4.4.3.4. Case 3 with $\alpha = -2$

Let $\boldsymbol{x}_{d,0}^{\infty}$ be defined by the limit for $\tau_L > 0$

$$\boldsymbol{x}_{d,0}^{\infty}=\lim_{\epsilon\to0}\boldsymbol{x}_{d}\left(t_{0}+\epsilon^{-2}\tau_{L}\right). \tag{4.288}$$

The outer equations for $\alpha = -2$ rely on the existence of the limit Eq. (4.288). The leading order equations are $2\left(n-p\right)+p$ algebraic equations,

$$\begin{aligned}\mathbf{0}=&\boldsymbol{\mathcal{Q}}_{\boldsymbol{\mathcal{S}}}^{T}\left(\boldsymbol{X}_{L}\right)\left(\nabla\boldsymbol{R}^{T}\left(\boldsymbol{X}_{L}\right)+\boldsymbol{\mathcal{U}}^{T}\left(\boldsymbol{X}_{L}\right)\right)\boldsymbol{\mathcal{Q}}_{\boldsymbol{\mathcal{S}}}^{T}\left(\boldsymbol{X}_{L}\right)\boldsymbol{\Lambda}_{L}\\&+\boldsymbol{\mathcal{Q}}_{\boldsymbol{\mathcal{S}}}^{T}\left(\boldsymbol{X}_{L}\right)\boldsymbol{\mathcal{S}}\boldsymbol{\mathcal{Q}}_{\boldsymbol{\mathcal{S}}}\left(\boldsymbol{X}_{L}\right)\left(\boldsymbol{X}_{L}-\boldsymbol{x}_{d,0}^{\infty}\right),\end{aligned} \tag{4.289}$$

$$\begin{aligned}\mathbf{0}=&\boldsymbol{\mathcal{P}}_{\boldsymbol{\mathcal{S}}}^{T}\left(\boldsymbol{X}_{L}\right)\left(\nabla\boldsymbol{R}^{T}\left(\boldsymbol{X}_{L}\right)+\boldsymbol{\mathcal{U}}^{T}\left(\boldsymbol{X}_{L}\right)\right)\boldsymbol{\mathcal{Q}}_{\boldsymbol{\mathcal{S}}}^{T}\left(\boldsymbol{X}_{L}\right)\boldsymbol{\Lambda}_{L}\\&+\boldsymbol{\mathcal{P}}_{\boldsymbol{\mathcal{S}}}^{T}\left(\boldsymbol{X}_{L}\right)\boldsymbol{\mathcal{S}}\boldsymbol{\mathcal{P}}_{\boldsymbol{\mathcal{S}}}\left(\boldsymbol{X}_{L}\right)\left(\boldsymbol{X}_{L}-\boldsymbol{x}_{d,0}^{\infty}\right),\end{aligned} \tag{4.290}$$

$$\mathbf{0}=\boldsymbol{\mathcal{Q}}_{\boldsymbol{\mathcal{S}}}\left(\boldsymbol{X}_{L}\right)\boldsymbol{R}\left(\boldsymbol{X}_{L}\right). \tag{4.291}$$

4.4.4. Inner equations - right side

The boundary layer at the right hand is similarly dealt with as the boundary layer at the left hand side. The new time scale is

$$\tau_{R}=\left(t_{1}-t\right)/\epsilon^{\alpha}, \tag{4.292}$$

which vanishes at $t = t_1$. The inner solutions are denoted by capital letters with an index R,

$$\boldsymbol{X}_R(\tau_R) = \boldsymbol{X}_R((t_1 - t)/\epsilon^\alpha) = \boldsymbol{x}(t), \qquad \boldsymbol{\Lambda}_R(\tau_R) = \boldsymbol{\Lambda}_R((t_1 - t)/\epsilon^\alpha) = \boldsymbol{\lambda}(t). \tag{4.293}$$

The terminal conditions Eq. (4.201) imply initial conditions for $\boldsymbol{X}_R(\tau_R)$ as

$$\boldsymbol{X}_R(0) = \boldsymbol{x}_1. \tag{4.294}$$

The determination of α by dominant balance and the derivation of the leading order equations proceeds analogous to the inner equations on the left side. The only difference is that a minus sign appears for time derivatives of odd order. Note that $\boldsymbol{\mathcal{V}}(\boldsymbol{x}, -\boldsymbol{y}) = -\boldsymbol{\mathcal{V}}(\boldsymbol{x}, \boldsymbol{y})$.

4.4.4.1. Case 1 with $\alpha = 2$

The leading order equations are $2(n-p)$ first order differential equations and p second order differential equations,

$$\boldsymbol{\mathcal{Q}}_{\boldsymbol{\mathcal{S}}}^T(\boldsymbol{X}_R)\boldsymbol{\Lambda}_R' = -\boldsymbol{\mathcal{Q}}_{\boldsymbol{\mathcal{S}}}^T(\boldsymbol{X}_R)\boldsymbol{\mathcal{V}}^T(\boldsymbol{X}_R, \boldsymbol{X}_R')\left(\boldsymbol{\Gamma}_{\boldsymbol{\mathcal{S}}}(\boldsymbol{X}_R)\boldsymbol{X}_R' + \boldsymbol{\mathcal{Q}}_{\boldsymbol{\mathcal{S}}}^T(\boldsymbol{X}_R)\boldsymbol{\Lambda}_R\right) \tag{4.295}$$

$$\begin{aligned}\frac{\partial}{\partial\tau_R}\left(\boldsymbol{\Gamma}_{\boldsymbol{\mathcal{S}}}(\boldsymbol{X}_R)\boldsymbol{X}_R'\right) = &-\boldsymbol{\mathcal{P}}_{\boldsymbol{\mathcal{S}}}^T(\boldsymbol{X}_R)\left(\boldsymbol{\mathcal{V}}^T(\boldsymbol{X}_R, \boldsymbol{X}_R') + \boldsymbol{\mathcal{Q}}_{\boldsymbol{\mathcal{S}}}^{T\prime}(\boldsymbol{X}_R)\right)\boldsymbol{\mathcal{Q}}_{\boldsymbol{\mathcal{S}}}^T(\boldsymbol{X}_R)\boldsymbol{\Lambda}_R \\ &- \boldsymbol{\mathcal{P}}_{\boldsymbol{\mathcal{S}}}^T(\boldsymbol{X}_R)\boldsymbol{\mathcal{V}}^T(\boldsymbol{X}_R, \boldsymbol{X}_R')\boldsymbol{\Gamma}_{\boldsymbol{\mathcal{S}}}(\boldsymbol{X}_R)\boldsymbol{X}_R', \end{aligned} \tag{4.296}$$

$$\boldsymbol{\mathcal{Q}}_{\boldsymbol{\mathcal{S}}}(\boldsymbol{X}_R)\boldsymbol{X}_R' = \boldsymbol{0}. \tag{4.297}$$

Eqs. (4.295)-(4.297) are not identical in form to their counterparts for the left boundary layer Eqs. (4.276)-(4.278).

4.4.4.2. Case 2.1 with $\alpha = 1$

As long as

$$\boldsymbol{\mathcal{P}}_{\boldsymbol{\mathcal{S}}}^T(\boldsymbol{X}_R)\left(\boldsymbol{\mathcal{V}}^T(\boldsymbol{X}_R, \boldsymbol{X}_R') + \boldsymbol{\mathcal{Q}}_{\boldsymbol{\mathcal{S}}}^{T\prime}(\boldsymbol{X}_R)\right)\boldsymbol{\mathcal{Q}}_{\boldsymbol{\mathcal{S}}}^T(\boldsymbol{X}_R)\boldsymbol{\Lambda}_R \neq \boldsymbol{0}, \tag{4.298}$$

the leading order equations are $2(n-p)+p$ first order differential equations

$$\boldsymbol{\mathcal{Q}}_{\boldsymbol{\mathcal{S}}}^T(\boldsymbol{X}_R)\boldsymbol{\Lambda}_R' = \boldsymbol{\mathcal{Q}}_{\boldsymbol{\mathcal{S}}}^T(\boldsymbol{X}_R)\boldsymbol{\mathcal{V}}^T(\boldsymbol{X}_R, \boldsymbol{X}_R')\boldsymbol{\mathcal{Q}}_{\boldsymbol{\mathcal{S}}}^T(\boldsymbol{X}_R)\boldsymbol{\Lambda}_R, \tag{4.299}$$

$$\boldsymbol{0} = \boldsymbol{\mathcal{P}}_{\boldsymbol{\mathcal{S}}}^T(\boldsymbol{X}_R)\left(\boldsymbol{\mathcal{V}}^T(\boldsymbol{X}_R, \boldsymbol{X}_R') + \boldsymbol{\mathcal{Q}}_{\boldsymbol{\mathcal{S}}}^{T\prime}(\boldsymbol{X}_R)\right)\boldsymbol{\mathcal{Q}}_{\boldsymbol{\mathcal{S}}}^T(\boldsymbol{X}_R)\boldsymbol{\Lambda}_R, \tag{4.300}$$

$$\boldsymbol{\mathcal{Q}}_{\boldsymbol{\mathcal{S}}}(\boldsymbol{X}_R)\boldsymbol{X}_R' = \boldsymbol{0}. \tag{4.301}$$

The leading order equations are identical in form to their counterparts for the left boundary layer.

4.4.4.3. Case 2.2 with $\alpha = 1$

The leading order equations for $\alpha = 1$ change if

$$\boldsymbol{\mathcal{P}}_{\boldsymbol{\mathcal{S}}}^{T}\left(\boldsymbol{X}_{R}\right)\left(\boldsymbol{\mathcal{V}}^{T}\left(\boldsymbol{X}_{R},\boldsymbol{X}_{R}^{\prime}\right)+\boldsymbol{\mathcal{Q}}_{\boldsymbol{\mathcal{S}}}^{T\prime}\left(\boldsymbol{X}_{R}\right)\right)\boldsymbol{\mathcal{Q}}_{\boldsymbol{\mathcal{S}}}^{T}\left(\boldsymbol{X}_{R}\right)\boldsymbol{\Lambda}_{R}=\mathbf{0}. \tag{4.302}$$

In this case, the leading order equations are $2\left(n-p\right)$ first order differential equations and p second order differential equations,

$$\boldsymbol{\mathcal{Q}}_{\boldsymbol{\mathcal{S}}}^{T}\left(\boldsymbol{X}_{R}\right)\boldsymbol{\Lambda}_{R}^{\prime}=-\boldsymbol{\mathcal{Q}}_{\boldsymbol{\mathcal{S}}}^{T}\left(\boldsymbol{X}_{R}\right)\boldsymbol{\mathcal{V}}^{T}\left(\boldsymbol{X}_{R},\boldsymbol{X}_{R}^{\prime}\right)\boldsymbol{\mathcal{Q}}_{\boldsymbol{\mathcal{S}}}^{T}\left(\boldsymbol{X}_{R}\right)\boldsymbol{\Lambda}_{R}, \tag{4.303}$$

$$\begin{aligned}\frac{\partial}{\partial\tau_{R}}\left(\boldsymbol{\Gamma}_{\boldsymbol{\mathcal{S}}}\left(\boldsymbol{X}_{R}\right)\boldsymbol{X}_{R}^{\prime}\right)&=-\boldsymbol{\mathcal{P}}_{\boldsymbol{\mathcal{S}}}^{T}\left(\boldsymbol{X}_{R}\right)\boldsymbol{\mathcal{V}}^{T}\left(\boldsymbol{X}_{R},\boldsymbol{X}_{R}^{\prime}\right)\boldsymbol{\Gamma}_{\boldsymbol{\mathcal{S}}}\left(\boldsymbol{X}_{R}\right)\boldsymbol{X}_{R}^{\prime}\\&\quad+\boldsymbol{\mathcal{P}}_{\boldsymbol{\mathcal{S}}}^{T}\left(\boldsymbol{X}_{R}\right)\left(\nabla\boldsymbol{R}^{T}\left(\boldsymbol{X}_{R}\right)+\boldsymbol{\mathcal{U}}^{T}\left(\boldsymbol{X}_{R}\right)\right)\boldsymbol{\mathcal{Q}}_{\boldsymbol{\mathcal{S}}}^{T}\left(\boldsymbol{X}_{R}\right)\boldsymbol{\Lambda}_{R}\\&\quad+\boldsymbol{\mathcal{P}}_{\boldsymbol{\mathcal{S}}}^{T}\left(\boldsymbol{X}_{R}\right)\boldsymbol{\mathcal{S}}\boldsymbol{\mathcal{P}}_{\boldsymbol{\mathcal{S}}}\left(\boldsymbol{X}_{R}\right)\left(\boldsymbol{X}_{R}-\boldsymbol{x}_{d}\left(t_{1}\right)\right),\end{aligned} \tag{4.304}$$

$$\boldsymbol{\mathcal{Q}}_{\boldsymbol{\mathcal{S}}}\left(\boldsymbol{X}_{R}\right)\boldsymbol{X}_{R}^{\prime}=\mathbf{0}. \tag{4.305}$$

The leading order equations are identical in form to their counterparts for the left boundary layer. This case corresponds to the left inner equations of the two-dimensional dynamical system from Section 4.1. Equations (4.303)-(4.305) reduce to the left inner equations derived in Section 4.1.5.2.

4.4.4.4. Case 3 with $\alpha = -2$

Let $\boldsymbol{x}_{d,1}^{\infty}$ be defined by the limit for $\tau_R > 0$

$$\boldsymbol{x}_{d,1}^{\infty}=\lim_{\epsilon\to0}\boldsymbol{x}_{d}\left(t_{1}-\epsilon^{-2}\tau_{R}\right). \tag{4.306}$$

The outer equations for $\alpha = -2$ rely on the existence of the limit Eq. (4.306). The leading order equations are $2\left(n-p\right)+p$ algebraic equations,

$$\begin{aligned}\mathbf{0}&=\boldsymbol{\mathcal{Q}}_{\boldsymbol{\mathcal{S}}}^{T}\left(\boldsymbol{X}_{R}\right)\left(\nabla\boldsymbol{R}^{T}\left(\boldsymbol{X}_{R}\right)+\boldsymbol{\mathcal{U}}^{T}\left(\boldsymbol{X}_{R}\right)\right)\boldsymbol{\mathcal{Q}}_{\boldsymbol{\mathcal{S}}}^{T}\left(\boldsymbol{X}_{R}\right)\boldsymbol{\Lambda}_{R}\\&\quad+\boldsymbol{\mathcal{Q}}_{\boldsymbol{\mathcal{S}}}^{T}\left(\boldsymbol{X}_{R}\right)\boldsymbol{\mathcal{S}}\boldsymbol{\mathcal{Q}}_{\boldsymbol{\mathcal{S}}}\left(\boldsymbol{X}_{R}\right)\left(\boldsymbol{X}_{R}-\boldsymbol{x}_{d,1}^{\infty}\right),\end{aligned} \tag{4.307}$$

$$\begin{aligned}\mathbf{0}&=\boldsymbol{\mathcal{P}}_{\boldsymbol{\mathcal{S}}}^{T}\left(\boldsymbol{X}_{R}\right)\left(\nabla\boldsymbol{R}^{T}\left(\boldsymbol{X}_{R}\right)+\boldsymbol{\mathcal{U}}^{T}\left(\boldsymbol{X}_{R}\right)\right)\boldsymbol{\mathcal{Q}}_{\boldsymbol{\mathcal{S}}}^{T}\left(\boldsymbol{X}_{R}\right)\boldsymbol{\Lambda}_{R}\\&\quad+\boldsymbol{\mathcal{P}}_{\boldsymbol{\mathcal{S}}}^{T}\left(\boldsymbol{X}_{R}\right)\boldsymbol{\mathcal{S}}\boldsymbol{\mathcal{P}}_{\boldsymbol{\mathcal{S}}}\left(\boldsymbol{X}_{R}\right)\left(\boldsymbol{X}_{R}-\boldsymbol{x}_{d,1}^{\infty}\right),\end{aligned} \tag{4.308}$$

$$\mathbf{0}=\boldsymbol{\mathcal{Q}}_{\boldsymbol{\mathcal{S}}}\left(\boldsymbol{X}_{R}\right)\boldsymbol{R}\left(\boldsymbol{X}_{R}\right). \tag{4.309}$$

These equations are identical in form to their counterparts for the left boundary layers.

4.4.5. Discussion of inner equations

Several cases of inner equations are possible for general dynamical systems. Different cases lead to different numbers of differential and algebraic equations. Consequently, the number of boundary conditions which can be accommodated by the inner equations differ from case to case. In the following, we focus on cases providing the maximum number of $2n$ boundary conditions. These are the cases with $2\,(n-p)$ first order and p second order differential equations given by Case 1, see Sections 4.4.3.1 and 4.4.4.1, and Case 2.2, see Sections 4.4.3.3 and 4.4.4.3. This choice is also motivated by the fact that Case 2.2 corresponds to the left and right inner equations of the two-dimensional dynamical system from Section 4.1. Both cases imply a constant state projection $\boldsymbol{\mathcal{Q}}_{\boldsymbol{\mathcal{S}}}\left(\boldsymbol{X}_{L/R}\right)\boldsymbol{X}_{L/R}$, see Eqs. (4.278) and (4.287). With the initial and terminal conditions for the state, Eqs. (4.271) and Eqs. (4.294), respectively, follows

$$\boldsymbol{\mathcal{Q}}_{\boldsymbol{\mathcal{S}}}\left(\boldsymbol{X}_L\left(\tau_L\right)\right)\boldsymbol{X}_L\left(\tau_L\right) = \boldsymbol{\mathcal{Q}}_{\boldsymbol{\mathcal{S}}}\left(\boldsymbol{X}_L\left(\tau_L\right)\right)\boldsymbol{x}_0, \tag{4.310}$$

$$\boldsymbol{\mathcal{Q}}_{\boldsymbol{\mathcal{S}}}\left(\boldsymbol{X}_R\left(\tau_R\right)\right)\boldsymbol{X}_R\left(\tau_R\right) = \boldsymbol{\mathcal{Q}}_{\boldsymbol{\mathcal{S}}}\left(\boldsymbol{X}_R\left(\tau_R\right)\right)\boldsymbol{x}_1. \tag{4.311}$$

In principle, all cases of inner equations listed in Sections 4.4.3 and 4.4.4 can play a role for the perturbative solution. Furthermore, more general scalings are possible for nonlinear state and co-state equations. The scaling might not only involve a rescaled time but also rescaled states and co-states as

$$\boldsymbol{X}_L\left(\tau_L\right) = \epsilon^{\beta}\boldsymbol{x}\left(t_0+\epsilon^{\alpha}\tau_L\right), \qquad \boldsymbol{\Lambda}_L\left(\tau_L\right) = \epsilon^{\gamma}\boldsymbol{x}\left(t_0+\epsilon^{\alpha}\tau_L\right). \tag{4.312}$$

Dominant balance arguments have to be applied to determine all possible combinations of exponents α, β, and γ. Usually, the values of β and γ depend explicitly on the form of all nonlinearities of the necessary optimality conditions. A larger variety of scalings can lead to much more difficult boundary layer structures than the simple boundary layers encountered for the two-dimensional system of Section 4.1. Multiple boundary layers are successions of boundary layers connecting the initial conditions with the outer solutions by two or more scaling regimes. Other possibilities are nested boundary layers, also called inner-inner boundary layers, or interior boundary layers located inside the time domain (Bender and Orszag, 2010). All cases of inner equations have to satisfy appropriate matching conditions connecting them to their neighboring inner or outer equations. A perturbative solution uniformly valid over the entire time interval is guaranteed only if a combination of inner and outer solutions satisfying their appropriate initial, terminal, and matching conditions exists. An exhaustive treatment including proofs for the existence and uniqueness of solutions for all possible combinations of scaling regimes is restricted to specific control systems and not performed here.

4.4.6. Matching

Here, the matching conditions for Case 1, see Sections 4.4.3.1 and 4.4.4.1, and Case 2.2, see Sections 4.4.3.3 and 4.4.4.3 are discussed. On the left side, the matching

conditions are

$$\lim_{\tau_L \to \infty} \boldsymbol{\mathcal{Q}}_{\boldsymbol{\mathcal{S}}}^T \left(\boldsymbol{X}_L \left(\tau_L \right) \right) \boldsymbol{\Lambda}_L \left(\tau_L \right) = \lim_{t \to t_0} \boldsymbol{\mathcal{Q}}_{\boldsymbol{\mathcal{S}}}^T \left(\boldsymbol{x}_O \left(t \right) \right) \boldsymbol{\lambda}_O \left(t \right), \tag{4.313}$$

$$\lim_{\tau_L \to \infty} \boldsymbol{\mathcal{Q}}_{\boldsymbol{\mathcal{S}}} \left(\boldsymbol{X}_L \left(\tau_L \right) \right) \boldsymbol{X}_L \left(\tau_L \right) = \lim_{t \to t_0} \boldsymbol{\mathcal{Q}}_{\boldsymbol{\mathcal{S}}} \left(\boldsymbol{x}_O \left(t \right) \right) \boldsymbol{x}_O \left(t \right), \tag{4.314}$$

$$\lim_{\tau_L \to \infty} \boldsymbol{\mathcal{P}}_{\boldsymbol{\mathcal{S}}} \left(\boldsymbol{X}_L \left(\tau_L \right) \right) \boldsymbol{X}_L \left(\tau_L \right) = \lim_{t \to t_0} \boldsymbol{\mathcal{P}}_{\boldsymbol{\mathcal{S}}} \left(\boldsymbol{x}_O \left(t \right) \right) \boldsymbol{x}_O \left(t \right). \tag{4.315}$$

Analogously, the matching conditions at the right side are

$$\lim_{\tau_R \to \infty} \boldsymbol{\mathcal{Q}}_{\boldsymbol{\mathcal{S}}}^T \left(\boldsymbol{X}_R \left(\tau_R \right) \right) \boldsymbol{\Lambda}_R \left(\tau_R \right) = \lim_{t \to t_1} \boldsymbol{\mathcal{Q}}_{\boldsymbol{\mathcal{S}}}^T \left(\boldsymbol{x}_O \left(t \right) \right) \boldsymbol{\lambda}_O \left(t \right), \tag{4.316}$$

$$\lim_{\tau_R \to \infty} \boldsymbol{\mathcal{Q}}_{\boldsymbol{\mathcal{S}}} \left(\boldsymbol{X}_R \left(\tau_R \right) \right) \boldsymbol{X}_R \left(\tau_R \right) = \lim_{t \to t_1} \boldsymbol{\mathcal{Q}}_{\boldsymbol{\mathcal{S}}} \left(\boldsymbol{x}_O \left(t \right) \right) \boldsymbol{x}_O \left(t \right), \tag{4.317}$$

$$\lim_{\tau_R \to \infty} \boldsymbol{\mathcal{P}}_{\boldsymbol{\mathcal{S}}} \left(\boldsymbol{X}_R \left(\tau_R \right) \right) \boldsymbol{X}_R \left(\tau_R \right) = \lim_{t \to t_1} \boldsymbol{\mathcal{P}}_{\boldsymbol{\mathcal{S}}} \left(\boldsymbol{x}_O \left(t \right) \right) \boldsymbol{x}_O \left(t \right). \tag{4.318}$$

Adding Eq. (4.314) and (4.315) yields

$$\lim_{\tau_L \to \infty} \boldsymbol{X}_L \left(\tau_L \right) = \boldsymbol{x}_O \left(t_0 \right), \tag{4.319}$$

and similarly for the right matching conditions

$$\lim_{\tau_R \to \infty} \boldsymbol{X}_R \left(\tau_R \right) = \boldsymbol{x}_O \left(t_1 \right). \tag{4.320}$$

The last two equations yield an identity for the projector,

$$\lim_{\tau_L \to \infty} \boldsymbol{\mathcal{Q}}_{\boldsymbol{\mathcal{S}}} \left(\boldsymbol{X}_L \left(\tau_L \right) \right) = \boldsymbol{\mathcal{Q}}_{\boldsymbol{\mathcal{S}}} \left(\boldsymbol{x}_O \left(t_0 \right) \right), \tag{4.321}$$

$$\lim_{\tau_R \to \infty} \boldsymbol{\mathcal{Q}}_{\boldsymbol{\mathcal{S}}} \left(\boldsymbol{X}_R \left(\tau_R \right) \right) = \boldsymbol{\mathcal{Q}}_{\boldsymbol{\mathcal{S}}} \left(\boldsymbol{x}_O \left(t_1 \right) \right), \tag{4.322}$$

and similarly for $\boldsymbol{\mathcal{P}}_{\boldsymbol{\mathcal{S}}}$.

The conditions Eqs. (4.313) and (4.316) yield the boundary conditions for the inner co-states $\boldsymbol{\mathcal{Q}}_{\boldsymbol{\mathcal{S}}}^T \left(\boldsymbol{X}_L \left(\tau_L \right) \right) \boldsymbol{\Lambda}_L \left(\tau_L \right)$ and $\boldsymbol{\mathcal{Q}}_{\boldsymbol{\mathcal{S}}}^T \left(\boldsymbol{X}_R \left(\tau_R \right) \right) \boldsymbol{\Lambda}_R \left(\tau_R \right)$, respectively.

Evaluating the algebraic outer Eq. (4.258) at the initial time t_0 yields

$$\begin{aligned}
&\boldsymbol{\mathcal{P}}_{\boldsymbol{\mathcal{S}}} \left(\boldsymbol{x}_O \left(t_0 \right) \right) \boldsymbol{x}_O \left(t_0 \right) = \boldsymbol{\mathcal{P}}_{\boldsymbol{\mathcal{S}}} \left(\boldsymbol{x}_O \left(t_0 \right) \right) \boldsymbol{x}_d \left(t_0 \right) \\
&- \boldsymbol{\Omega}_{\boldsymbol{\mathcal{S}}} \left(\boldsymbol{x}_O \left(t_0 \right) \right) \left(\dot{\boldsymbol{\mathcal{Q}}}_{\boldsymbol{\mathcal{S}}}^T \left(\boldsymbol{x}_O \left(t_0 \right) \right) + \nabla \boldsymbol{R}^T \left(\boldsymbol{x}_O \left(t_0 \right) \right) \right) \boldsymbol{\mathcal{Q}}_{\boldsymbol{\mathcal{S}}}^T \left(\boldsymbol{x}_O \left(t_0 \right) \right) \boldsymbol{\lambda}_O \left(t_0 \right) \\
&- \boldsymbol{\Omega}_{\boldsymbol{\mathcal{S}}} \left(\boldsymbol{x}_O \left(t_0 \right) \right) \boldsymbol{\mathcal{W}}^T \left(\boldsymbol{x}_O \left(t_0 \right), \dot{\boldsymbol{x}}_O \left(t_0 \right) \right) \boldsymbol{\mathcal{Q}}_{\boldsymbol{\mathcal{S}}}^T \left(\boldsymbol{x}_O \left(t_0 \right) \right) \boldsymbol{\lambda}_O \left(t_0 \right).
\end{aligned} \tag{4.323}$$

Together with the matching condition Eq. (4.315), Eq. (4.323) results in an additional boundary condition for $\boldsymbol{\mathcal{P}}_{\boldsymbol{\mathcal{S}}} \left(\boldsymbol{X}_L \right) \boldsymbol{X}_L$,

$$\begin{aligned}
&\lim_{\tau_L \to \infty} \boldsymbol{\mathcal{P}}_{\boldsymbol{\mathcal{S}}} \left(\boldsymbol{x}_O \left(t_0 \right) \right) \boldsymbol{X}_L \left(\tau_L \right) = \boldsymbol{\mathcal{P}}_{\boldsymbol{\mathcal{S}}} \left(\boldsymbol{x}_O \left(t_0 \right) \right) \boldsymbol{x}_d \left(t_0 \right) \\
&- \boldsymbol{\Omega}_{\boldsymbol{\mathcal{S}}} \left(\boldsymbol{x}_O \left(t_0 \right) \right) \left(\dot{\boldsymbol{\mathcal{Q}}}_{\boldsymbol{\mathcal{S}}}^T \left(\boldsymbol{x}_O \left(t_0 \right) \right) + \nabla \boldsymbol{R}^T \left(\boldsymbol{x}_O \left(t_0 \right) \right) \right) \boldsymbol{\mathcal{Q}}_{\boldsymbol{\mathcal{S}}}^T \left(\boldsymbol{x}_O \left(t_0 \right) \right) \boldsymbol{\lambda}_O \left(t_0 \right) \\
&- \boldsymbol{\Omega}_{\boldsymbol{\mathcal{S}}} \left(\boldsymbol{x}_O \left(t_0 \right) \right) \boldsymbol{\mathcal{W}}^T \left(\boldsymbol{x}_O \left(t_0 \right), \dot{\boldsymbol{x}}_O \left(t_0 \right) \right) \boldsymbol{\mathcal{Q}}_{\boldsymbol{\mathcal{S}}}^T \left(\boldsymbol{x}_O \left(t_0 \right) \right) \boldsymbol{\lambda}_O \left(t_0 \right).
\end{aligned} \tag{4.324}$$

The existence of this limit, together with the result that $\boldsymbol{\mathcal{Q}}_{\boldsymbol{\mathcal{S}}}\left(\boldsymbol{X}_L\right)\boldsymbol{X}_L$ is constant, see Eq. (4.310), implies

$$\lim_{\tau_L\to\infty}\boldsymbol{X}_L'\left(\tau_L\right)=\boldsymbol{0}. \tag{4.325}$$

On the other hand, evaluating the algebraic outer Eq. (4.258) at the terminal time t_1 yields

$$\begin{aligned}&\boldsymbol{\mathcal{P}}_{\boldsymbol{\mathcal{S}}}\left(\boldsymbol{x}_O\left(t_1\right)\right)\boldsymbol{x}_O\left(t_1\right)=\boldsymbol{\mathcal{P}}_{\boldsymbol{\mathcal{S}}}\left(\boldsymbol{x}_O\left(t_1\right)\right)\boldsymbol{x}_d\left(t_1\right)\\&-\boldsymbol{\Omega}_{\boldsymbol{\mathcal{S}}}\left(\boldsymbol{x}_O\left(t_1\right)\right)\left(\dot{\boldsymbol{\mathcal{Q}}}_{\boldsymbol{\mathcal{S}}}^T\left(\boldsymbol{x}_O\left(t_1\right)\right)+\nabla\boldsymbol{R}^T\left(\boldsymbol{x}_O\left(t_1\right)\right)\right)\boldsymbol{\mathcal{Q}}_{\boldsymbol{\mathcal{S}}}^T\left(\boldsymbol{x}_O\left(t_1\right)\right)\boldsymbol{\lambda}_O\left(t_1\right)\\&-\boldsymbol{\Omega}_{\boldsymbol{\mathcal{S}}}\left(\boldsymbol{x}_O\left(t_1\right)\right)\boldsymbol{\mathcal{W}}^T\left(\boldsymbol{x}_O\left(t_1\right),\dot{\boldsymbol{x}}_O\left(t_1\right)\right)\boldsymbol{\mathcal{Q}}_{\boldsymbol{\mathcal{S}}}^T\left(\boldsymbol{x}_O\left(t_1\right)\right)\boldsymbol{\lambda}_O\left(t_1\right).\end{aligned} \tag{4.326}$$

Together with the matching condition Eq. (4.318), Eq. (4.326) results in an additional boundary condition for $\boldsymbol{\mathcal{P}}_{\boldsymbol{\mathcal{S}}}\left(\boldsymbol{X}_R\right)\boldsymbol{X}_R$ in the form

$$\begin{aligned}&\lim_{\tau_R\to\infty}\boldsymbol{\mathcal{P}}_{\boldsymbol{\mathcal{S}}}\left(\boldsymbol{X}_R\left(\tau_R\right)\right)\boldsymbol{X}_R\left(\tau_R\right)=\boldsymbol{\mathcal{P}}_{\boldsymbol{\mathcal{S}}}\left(\boldsymbol{x}_O\left(t_1\right)\right)\boldsymbol{x}_d\left(t_1\right)\\&-\boldsymbol{\Omega}_{\boldsymbol{\mathcal{S}}}\left(\boldsymbol{x}_O\left(t_1\right)\right)\left(\dot{\boldsymbol{\mathcal{Q}}}_{\boldsymbol{\mathcal{S}}}^T\left(\boldsymbol{x}_O\left(t_1\right)\right)+\nabla\boldsymbol{R}^T\left(\boldsymbol{x}_O\left(t_1\right)\right)\right)\boldsymbol{\mathcal{Q}}_{\boldsymbol{\mathcal{S}}}^T\left(\boldsymbol{x}_O\left(t_1\right)\right)\boldsymbol{\lambda}_O\left(t_1\right)\\&-\boldsymbol{\Omega}_{\boldsymbol{\mathcal{S}}}\left(\boldsymbol{x}_O\left(t_1\right)\right)\boldsymbol{\mathcal{W}}^T\left(\boldsymbol{x}_O\left(t_1\right),\dot{\boldsymbol{x}}_O\left(t_1\right)\right)\boldsymbol{\mathcal{Q}}_{\boldsymbol{\mathcal{S}}}^T\left(\boldsymbol{x}_O\left(t_1\right)\right)\boldsymbol{\lambda}_O\left(t_1\right).\end{aligned} \tag{4.327}$$

Similar to above, the existence of this limit, together with the result that $\boldsymbol{\mathcal{Q}}_{\boldsymbol{\mathcal{S}}}\left(\boldsymbol{X}_R\right)\boldsymbol{X}_R$ is constant, see Eq. (4.311), implies

$$\lim_{\tau_R\to\infty}\boldsymbol{X}_R'\left(\tau_R\right)=\boldsymbol{0}. \tag{4.328}$$

Finally, two matching conditions Eqs. (4.314) and (4.317) remain. Because of the constancy of $\boldsymbol{\mathcal{Q}}_{\boldsymbol{\mathcal{S}}}\left(\boldsymbol{X}_{L/R}\right)\boldsymbol{X}_{L/R}$, Eqs. (4.310) and (4.311), together with Eqs. (4.319) and (4.320), respectively, these can be written as

$$\boldsymbol{\mathcal{Q}}_{\boldsymbol{\mathcal{S}}}\left(\boldsymbol{x}_O\left(t_0\right)\right)\boldsymbol{x}_0=\boldsymbol{\mathcal{Q}}_{\boldsymbol{\mathcal{S}}}\left(\boldsymbol{x}_O\left(t_0\right)\right)\boldsymbol{x}_O\left(t_0\right), \tag{4.329}$$

$$\boldsymbol{\mathcal{Q}}_{\boldsymbol{\mathcal{S}}}\left(\boldsymbol{x}_O\left(t_1\right)\right)\boldsymbol{x}_1=\boldsymbol{\mathcal{Q}}_{\boldsymbol{\mathcal{S}}}\left(\boldsymbol{x}_O\left(t_1\right)\right)\boldsymbol{x}_O\left(t_1\right). \tag{4.330}$$

These are the boundary conditions for the outer equations, Eqs. (4.255) and (4.257). They depend only on the initial and terminal conditions $\boldsymbol{x}_0$ and $\boldsymbol{x}_1$, respectively. Hence, the solutions to the outer equations are independent of any details of the inner equations. In particular, the outer solutions are identical for both cases, Case 1 and Case 2.2, of inner equations discussed here.

Finally, it is possible to formally write down the composite solutions for the problem. The parts $\boldsymbol{\mathcal{Q}}_{\boldsymbol{\mathcal{S}}}\left(\boldsymbol{x}_{\text{comp}}\left(t\right)\right)\boldsymbol{x}_{\text{comp}}\left(t\right)$ and $\boldsymbol{\mathcal{Q}}_{\boldsymbol{\mathcal{S}}}\left(\boldsymbol{\lambda}_{\text{comp}}\left(t\right)\right)\boldsymbol{\lambda}_{\text{comp}}\left(t\right)$ do not exhibit boundary layers and are simply given by the solution to the outer equations,

$$\boldsymbol{\mathcal{Q}}_{\boldsymbol{\mathcal{S}}}\left(\boldsymbol{x}_{\text{comp}}\left(t\right)\right)\boldsymbol{x}_{\text{comp}}\left(t\right)=\boldsymbol{\mathcal{Q}}_{\boldsymbol{\mathcal{S}}}\left(\boldsymbol{x}_O\left(t\right)\right)\boldsymbol{x}_O\left(t\right), \tag{4.331}$$

$$\boldsymbol{\mathcal{Q}}_{\boldsymbol{\mathcal{S}}}\left(\boldsymbol{x}_{\text{comp}}\left(t\right)\right)\boldsymbol{\lambda}_{\text{comp}}\left(t\right)=\boldsymbol{\mathcal{Q}}_{\boldsymbol{\mathcal{S}}}\left(\boldsymbol{x}_O\left(t\right)\right)\boldsymbol{\lambda}_O\left(t\right). \tag{4.332}$$

The part $\boldsymbol{\mathcal{P}}_{\boldsymbol{\mathcal{S}}}\left(\boldsymbol{x}_{\text{comp}}\left(t\right)\right)\boldsymbol{x}_{\text{comp}}\left(t\right)$ contains boundary layers and is given by the sum of outer, left inner and right inner solution minus the overlaps $\boldsymbol{\mathcal{P}}_{\boldsymbol{\mathcal{S}}}\left(\boldsymbol{x}_O\left(t_0\right)\right)\boldsymbol{x}_O\left(t_0\right)$ and $\boldsymbol{\mathcal{P}}_{\boldsymbol{\mathcal{S}}}\left(\boldsymbol{x}_O\left(t_1\right)\right)\boldsymbol{x}_O\left(t_1\right)$,

$$\begin{aligned}
&\boldsymbol{\mathcal{P}}_{\boldsymbol{\mathcal{S}}}\left(\boldsymbol{x}_{\text{comp}}\left(t\right)\right)\boldsymbol{x}_{\text{comp}}\left(t\right) = \boldsymbol{\mathcal{P}}_{\boldsymbol{\mathcal{S}}}\left(\boldsymbol{x}_O\left(t\right)\right)\boldsymbol{x}_O\left(t\right)\\
&+\boldsymbol{\mathcal{P}}_{\boldsymbol{\mathcal{S}}}\left(\boldsymbol{X}_L\left(\epsilon^{-\alpha}\left(t-t_0\right)\right)\right)\boldsymbol{X}_L\left(\epsilon^{-\alpha}\left(t-t_0\right)\right) - \boldsymbol{\mathcal{P}}_{\boldsymbol{\mathcal{S}}}\left(\boldsymbol{x}_O\left(t_0\right)\right)\boldsymbol{x}_O\left(t_0\right)\\
&+\boldsymbol{\mathcal{P}}_{\boldsymbol{\mathcal{S}}}\left(\boldsymbol{X}_R\left(\epsilon^{-\alpha}\left(t_1-t\right)\right)\right)\boldsymbol{X}_R\left(\epsilon^{-\alpha}\left(t_1-t\right)\right) - \boldsymbol{\mathcal{P}}_{\boldsymbol{\mathcal{S}}}\left(\boldsymbol{x}_O\left(t_1\right)\right)\boldsymbol{x}_O\left(t_1\right). \qquad (4.333)
\end{aligned}$$

Finally, the controlled state reads as

$$\begin{aligned}
\boldsymbol{x}_{\text{comp}}\left(t\right) &= \boldsymbol{\mathcal{P}}_{\boldsymbol{\mathcal{S}}}\left(\boldsymbol{x}_{\text{comp}}\left(t\right)\right)\boldsymbol{x}_{\text{comp}}\left(t\right) + \boldsymbol{\mathcal{Q}}_{\boldsymbol{\mathcal{S}}}\left(\boldsymbol{x}_{\text{comp}}\left(t\right)\right)\boldsymbol{x}_{\text{comp}}\left(t\right)\\
&= \boldsymbol{x}_O\left(t\right) - \boldsymbol{\mathcal{P}}_{\boldsymbol{\mathcal{S}}}\left(\boldsymbol{x}_O\left(t_0\right)\right)\boldsymbol{x}_O\left(t_0\right) - \boldsymbol{\mathcal{P}}_{\boldsymbol{\mathcal{S}}}\left(\boldsymbol{x}_O\left(t_1\right)\right)\boldsymbol{x}_O\left(t_1\right)\\
&+\boldsymbol{\mathcal{P}}_{\boldsymbol{\mathcal{S}}}\left(\boldsymbol{X}_L\left(\epsilon^{-\alpha}\left(t-t_0\right)\right)\right)\boldsymbol{X}_L\left(\epsilon^{-\alpha}\left(t-t_0\right)\right)\\
&+\boldsymbol{\mathcal{P}}_{\boldsymbol{\mathcal{S}}}\left(\boldsymbol{X}_R\left(\epsilon^{-\alpha}\left(t_1-t\right)\right)\right)\boldsymbol{X}_R\left(\epsilon^{-\alpha}\left(t_1-t\right)\right). \qquad (4.334)
\end{aligned}$$

The composite control signal is given in terms of the composite solutions as

$$\boldsymbol{u}_{\text{comp}}\left(t\right) = \boldsymbol{\mathcal{B}}^g_{\boldsymbol{\mathcal{S}}}\left(\boldsymbol{x}_{\text{comp}}\left(t\right)\right)\left(\dot{\boldsymbol{x}}_{\text{comp}}\left(t\right) - \boldsymbol{R}\left(\boldsymbol{x}_{\text{comp}}\left(t\right)\right)\right). \qquad (4.335)$$

4.4.7. Exact state solution for $\epsilon = 0$

For a vanishing value of the regularization parameter ϵ, the inner solutions degenerate to jumps located at the time domain boundaries. The exact solution for the controlled state trajectory $\boldsymbol{x}\left(t\right)$ is entirely determined by the outer equations supplemented with appropriate boundary conditions and jumps. The time evolution of the state $\boldsymbol{x}\left(t\right)$ and co-state $\boldsymbol{\lambda}\left(t\right)$ is governed by $2\left(n-p\right)$ first order differential equations and $2p$ algebraic equations.

First, the parts $\boldsymbol{\mathcal{P}}^T_{\boldsymbol{\mathcal{S}}}\left(\boldsymbol{x}\right)\boldsymbol{\lambda}$ and $\boldsymbol{\mathcal{P}}_{\boldsymbol{\mathcal{S}}}\left(\boldsymbol{x}\right)\boldsymbol{x}$ are given by algebraic equations. The part $\boldsymbol{\mathcal{P}}^T_{\boldsymbol{\mathcal{S}}}\left(\boldsymbol{x}\right)\boldsymbol{\lambda}$ vanishes identically for all times,

$$\boldsymbol{\mathcal{P}}^T_{\boldsymbol{\mathcal{S}}}\left(\boldsymbol{x}\left(t\right)\right)\boldsymbol{\lambda}\left(t\right) = \boldsymbol{0}. \qquad (4.336)$$

The part $\boldsymbol{\mathcal{P}}_{\boldsymbol{\mathcal{S}}}\left(\boldsymbol{x}\right)\boldsymbol{x}$ behaves discontinuously at the domain boundaries,

$$\begin{aligned}
\boldsymbol{\mathcal{P}}_{\boldsymbol{\mathcal{S}}}\left(\boldsymbol{x}\left(t\right)\right)\boldsymbol{x}\left(t\right) &= \lim_{\epsilon\to 0}\boldsymbol{\mathcal{P}}_{\boldsymbol{\mathcal{S}}}\left(\boldsymbol{x}_{\text{comp}}\left(t\right)\right)\boldsymbol{x}_{\text{comp}}\left(t\right)\\
&= \begin{cases}
\boldsymbol{\mathcal{P}}_{\boldsymbol{\mathcal{S}}}\left(\boldsymbol{x}_0\right)\boldsymbol{x}_0, & t = t_0,\\
\boldsymbol{\mathcal{P}}_{\boldsymbol{\mathcal{S}}}\left(\boldsymbol{x}_O\left(t\right)\right)\boldsymbol{x}_O\left(t\right), & t_0 < t < t_1,\\
\boldsymbol{\mathcal{P}}_{\boldsymbol{\mathcal{S}}}\left(\boldsymbol{x}_1\right)\boldsymbol{x}_1, & t = t_1.
\end{cases} \qquad (4.337)
\end{aligned}$$

Inside the time domain, $\boldsymbol{\mathcal{P}}_{\boldsymbol{\mathcal{S}}}\left(\boldsymbol{x}\right)\boldsymbol{x}$ behaves continuously and is given in terms of $\boldsymbol{\mathcal{Q}}^T_{\boldsymbol{\mathcal{S}}}\left(\boldsymbol{x}_O\right)\boldsymbol{\lambda}_O$ and $\boldsymbol{\mathcal{Q}}_{\boldsymbol{\mathcal{S}}}\left(\boldsymbol{x}_O\right)\boldsymbol{x}_O$ as

$$\begin{aligned}
&\boldsymbol{\mathcal{P}}_{\boldsymbol{\mathcal{S}}}\left(\boldsymbol{x}_O\left(t\right)\right)\boldsymbol{x}_O\left(t\right) = \boldsymbol{\mathcal{P}}_{\boldsymbol{\mathcal{S}}}\left(\boldsymbol{x}_O\left(t\right)\right)\boldsymbol{x}_d\left(t\right)\\
&-\boldsymbol{\Omega}_{\boldsymbol{\mathcal{S}}}\left(\boldsymbol{x}_O\left(t\right)\right)\left(\dot{\boldsymbol{\mathcal{Q}}}^T_{\boldsymbol{\mathcal{S}}}\left(\boldsymbol{x}_O\left(t\right)\right) + \nabla\boldsymbol{R}^T\left(\boldsymbol{x}_O\left(t\right)\right)\right)\boldsymbol{\mathcal{Q}}^T_{\boldsymbol{\mathcal{S}}}\left(\boldsymbol{x}_O\left(t\right)\right)\boldsymbol{\lambda}_O\left(t\right)\\
&-\boldsymbol{\Omega}_{\boldsymbol{\mathcal{S}}}\left(\boldsymbol{x}_O\left(t\right)\right)\boldsymbol{\mathcal{W}}^T\left(\boldsymbol{x}_O\left(t\right),\dot{\boldsymbol{x}}_O\left(t\right)\right)\boldsymbol{\mathcal{Q}}^T_{\boldsymbol{\mathcal{S}}}\left(\boldsymbol{x}_O\left(t\right)\right)\boldsymbol{\lambda}_O\left(t\right). \qquad (4.338)
\end{aligned}$$

The parts $\boldsymbol{\mathcal{Q}}_{\boldsymbol{\mathcal{S}}}^{T}(\boldsymbol{x})\boldsymbol{\lambda}$ and $\boldsymbol{\mathcal{Q}}_{\boldsymbol{\mathcal{S}}}(\boldsymbol{x})\boldsymbol{x}$ are given as the solution to the outer equations

$$\boldsymbol{\mathcal{Q}}_{\boldsymbol{\mathcal{S}}}^{T}(\boldsymbol{x}(t))\boldsymbol{\lambda}(t) = \boldsymbol{\mathcal{Q}}_{\boldsymbol{\mathcal{S}}}^{T}(\boldsymbol{x}_O(t))\boldsymbol{\lambda}_O(t), \tag{4.339}$$

$$\boldsymbol{\mathcal{Q}}_{\boldsymbol{\mathcal{S}}}(\boldsymbol{x}(t))\boldsymbol{x}(t) = \boldsymbol{\mathcal{Q}}_{\boldsymbol{\mathcal{S}}}(\boldsymbol{x}_O(t))\boldsymbol{x}_O(t), \tag{4.340}$$

which satisfy

$$\begin{aligned}-\boldsymbol{\mathcal{Q}}_{\boldsymbol{\mathcal{S}}}^{T}(\boldsymbol{x}_O(t))\dot{\boldsymbol{\lambda}}_O(t) &= \boldsymbol{\mathcal{Q}}_{\boldsymbol{\mathcal{S}}}^{T}(\boldsymbol{x}_O(t))\nabla\boldsymbol{R}^{T}(\boldsymbol{x}_O(t))\boldsymbol{\mathcal{Q}}_{\boldsymbol{\mathcal{S}}}^{T}(\boldsymbol{x}_O(t))\boldsymbol{\lambda}_O(t)\\ &\quad+\boldsymbol{\mathcal{Q}}_{\boldsymbol{\mathcal{S}}}^{T}(\boldsymbol{x}_O(t))\boldsymbol{\mathcal{W}}^{T}(\boldsymbol{x}_O(t),\dot{\boldsymbol{x}}_O(t))\boldsymbol{\mathcal{Q}}_{\boldsymbol{\mathcal{S}}}^{T}(\boldsymbol{x}_O(t))\boldsymbol{\lambda}_O(t)\\ &\quad+\boldsymbol{\mathcal{Q}}_{\boldsymbol{\mathcal{S}}}^{T}(\boldsymbol{x}_O(t))\boldsymbol{\mathcal{S}}\boldsymbol{\mathcal{Q}}_{\boldsymbol{\mathcal{S}}}(\boldsymbol{x}_O(t))(\boldsymbol{x}_O(t)-\boldsymbol{x}_d(t)), \qquad (4.341)\\ \boldsymbol{\mathcal{Q}}_{\boldsymbol{\mathcal{S}}}(\boldsymbol{x}_O(t))\dot{\boldsymbol{x}}_O(t) &= \boldsymbol{\mathcal{Q}}_{\boldsymbol{\mathcal{S}}}(\boldsymbol{x}_O(t))\boldsymbol{R}(\boldsymbol{x}_O(t)). \qquad (4.342)\end{aligned}$$

These equations have to satisfy the boundary conditions

$$\boldsymbol{\mathcal{Q}}_{\boldsymbol{\mathcal{S}}}(\boldsymbol{x}_O(t_0))\boldsymbol{x}_0 = \boldsymbol{\mathcal{Q}}_{\boldsymbol{\mathcal{S}}}(\boldsymbol{x}_O(t_0))\boldsymbol{x}_O(t_0), \tag{4.343}$$

$$\boldsymbol{\mathcal{Q}}_{\boldsymbol{\mathcal{S}}}(\boldsymbol{x}_O(t_1))\boldsymbol{x}_1 = \boldsymbol{\mathcal{Q}}_{\boldsymbol{\mathcal{S}}}(\boldsymbol{x}_O(t_1))\boldsymbol{x}_O(t_1). \tag{4.344}$$

The full state can be expressed as

$$\boldsymbol{x}(t) = \lim_{\epsilon\to 0}\boldsymbol{x}_{\mathrm{comp}}(t) = \begin{cases}\boldsymbol{x}_0, & t = t_0,\\ \boldsymbol{x}_O(t), & t_0 < t < t_1,\\ \boldsymbol{x}_1, & t = t_1.\end{cases} \tag{4.345}$$

The jumps exhibited by $\boldsymbol{\mathcal{P}}_{\boldsymbol{\mathcal{S}}}(\boldsymbol{x})\boldsymbol{x}$ at the beginning and the end of the time domain are remnants of the boundary layers. Together with Eq. (4.338), their heights are given by

$$\begin{aligned}&\boldsymbol{\mathcal{P}}_{\boldsymbol{\mathcal{S}}}(\boldsymbol{x}_O(t_0))\boldsymbol{x}_O(t_0) - \boldsymbol{\mathcal{P}}_{\boldsymbol{\mathcal{S}}}(\boldsymbol{x}_0)\boldsymbol{x}_0 = \boldsymbol{\mathcal{P}}_{\boldsymbol{\mathcal{S}}}(\boldsymbol{x}_O(t_0))\boldsymbol{x}_d(t_0) - \boldsymbol{\mathcal{P}}_{\boldsymbol{\mathcal{S}}}(\boldsymbol{x}_0)\boldsymbol{x}_0\\ &-\boldsymbol{\Omega}_{\boldsymbol{\mathcal{S}}}(\boldsymbol{x}_O(t_0))\left(\dot{\boldsymbol{\mathcal{Q}}}_{\boldsymbol{\mathcal{S}}}^{T}(\boldsymbol{x}_O(t_0)) + \nabla\boldsymbol{R}^{T}(\boldsymbol{x}_O(t_0))\right)\boldsymbol{\mathcal{Q}}_{\boldsymbol{\mathcal{S}}}^{T}(\boldsymbol{x}_O(t_0))\boldsymbol{\lambda}_O(t_0)\\ &-\boldsymbol{\Omega}_{\boldsymbol{\mathcal{S}}}(\boldsymbol{x}_O(t_0))\boldsymbol{\mathcal{W}}^{T}(\boldsymbol{x}_O(t_0),\dot{\boldsymbol{x}}_O(t_0))\boldsymbol{\mathcal{Q}}_{\boldsymbol{\mathcal{S}}}^{T}(\boldsymbol{x}_O(t_0))\boldsymbol{\lambda}_O(t_0), \qquad (4.346)\end{aligned}$$

and

$$\begin{aligned}&\boldsymbol{\mathcal{P}}_{\boldsymbol{\mathcal{S}}}(\boldsymbol{x}_O(t_1))\boldsymbol{x}_O(t_1) - \boldsymbol{\mathcal{P}}_{\boldsymbol{\mathcal{S}}}(\boldsymbol{x}_1)\boldsymbol{x}_1 = \boldsymbol{\mathcal{P}}_{\boldsymbol{\mathcal{S}}}(\boldsymbol{x}_O(t_1))\boldsymbol{x}_d(t_1) - \boldsymbol{\mathcal{P}}_{\boldsymbol{\mathcal{S}}}(\boldsymbol{x}_1)\boldsymbol{x}_1\\ &-\boldsymbol{\Omega}_{\boldsymbol{\mathcal{S}}}(\boldsymbol{x}_O(t_1))\left(\dot{\boldsymbol{\mathcal{Q}}}_{\boldsymbol{\mathcal{S}}}^{T}(\boldsymbol{x}_O(t_1)) + \nabla\boldsymbol{R}^{T}(\boldsymbol{x}_O(t_1))\right)\boldsymbol{\mathcal{Q}}_{\boldsymbol{\mathcal{S}}}^{T}(\boldsymbol{x}_O(t_1))\boldsymbol{\lambda}_O(t_1)\\ &-\boldsymbol{\Omega}_{\boldsymbol{\mathcal{S}}}(\boldsymbol{x}_O(t_1))\boldsymbol{\mathcal{W}}^{T}(\boldsymbol{x}_O(t_1),\dot{\boldsymbol{x}}_O(t_1))\boldsymbol{\mathcal{Q}}_{\boldsymbol{\mathcal{S}}}^{T}(\boldsymbol{x}_O(t_1))\boldsymbol{\lambda}_O(t_1), \qquad (4.347)\end{aligned}$$

respectively.

In conclusion, the jump heights are entirely determined in terms of the solutions to the outer equation together with the initial and terminal conditions for the state. Thus, for $\epsilon = 0$, no traces of the boundary layers survive except their mere existence

and location. The inner equations play no role for the *form* of the exact state and costate trajectory for $\epsilon = 0$. However, the inner equations play a role for the *existence* of the exact solution. The existence of jumps of appropriate height can only be guaranteed if inner solutions satisfying appropriate initial, terminal, and matching conditions exist.

Here, we assumed that solutions exist for the inner equations either given by Case 1, see Sections 4.4.3.1 and 4.4.4.1, or Case 2.2, see Sections 4.4.3.3 and 4.4.4.3. If other scaling regimes not given by Case 1 or Case 2.2 play a role, the initial conditions Eqs. (4.343) and (4.344) for the outer equations might change.

We emphasize that the exact solution stated in this section is highly formal. The expressions for the parts $\boldsymbol{\mathcal{P}}_{\boldsymbol{\mathcal{S}}}(\boldsymbol{x}_O)\,\boldsymbol{x}_O$ and $\boldsymbol{\mathcal{Q}}_{\boldsymbol{\mathcal{S}}}(\boldsymbol{x}_O)\,\boldsymbol{x}_O$ are not closed form expressions as long as also the projectors depend on $\boldsymbol{x}_O$. In general, the expression Eq. (4.338) for $\boldsymbol{\mathcal{P}}_{\boldsymbol{\mathcal{S}}}(\boldsymbol{x}_O)\,\boldsymbol{x}_O$ is a nonlinear equation for $\boldsymbol{\mathcal{P}}_{\boldsymbol{\mathcal{S}}}(\boldsymbol{x}_O)\,\boldsymbol{x}_O$. Closed form expressions can be obtained by transforming the state $\boldsymbol{x}$ such that the projectors $\boldsymbol{\mathcal{P}}_{\boldsymbol{\mathcal{S}}}$ and $\boldsymbol{\mathcal{Q}}_{\boldsymbol{\mathcal{S}}}$ are diagonal.

4.4.8. Exact control solution for $\epsilon = 0$

Formally, the control signal is given in terms of the controlled state trajectory by the expression

$$\boldsymbol{u}(t) = \boldsymbol{\mathcal{B}}^{g}_{\boldsymbol{\mathcal{S}}}(\boldsymbol{x}(t))\left(\dot{\boldsymbol{x}}(t) - \boldsymbol{R}(\boldsymbol{x}(t))\right). \tag{4.348}$$

However, care has to be taken when evaluating the time derivative $\dot{\boldsymbol{x}}(t)$ at the time domain boundaries. To determine the control signal at these points, it is necessary to analyze the expression

$$\boldsymbol{u}_{\text{comp}}(t) = \boldsymbol{\mathcal{B}}^{g}_{\boldsymbol{\mathcal{S}}}(\boldsymbol{x}_{\text{comp}}(t))\left(\dot{\boldsymbol{x}}_{\text{comp}}(t) - \boldsymbol{R}(\boldsymbol{x}_{\text{comp}}(t))\right) \tag{4.349}$$

in the limit $\epsilon \to 0$. All terms except $\dot{\boldsymbol{x}}_{\text{comp}}(t)$ are well behaved. The term $\dot{\boldsymbol{x}}_{\text{comp}}(t)$ requires the investigation of the limit

$$\begin{aligned}\lim_{\epsilon\to 0}\dot{\boldsymbol{x}}_{\text{comp}}(t) = \dot{\boldsymbol{x}}_O(t)\\ &+ \lim_{\epsilon\to 0}\frac{d}{dt}\left(\boldsymbol{\mathcal{P}}_{\boldsymbol{\mathcal{S}}}\left(\boldsymbol{X}_L\left(\epsilon^{-\alpha}(t-t_0)\right)\right)\boldsymbol{X}_L\left(\epsilon^{-\alpha}(t-t_0)\right)\right)\\ &+ \lim_{\epsilon\to 0}\frac{d}{dt}\left(\boldsymbol{\mathcal{P}}_{\boldsymbol{\mathcal{S}}}\left(\boldsymbol{X}_R\left(\epsilon^{-\alpha}(t_1-t)\right)\right)\boldsymbol{X}_R\left(\epsilon^{-\alpha}(t_1-t)\right)\right).\end{aligned} \tag{4.350}$$

Similar as for two-dimensional dynamical systems in Section 4.1, it is possible to prove that $\frac{d}{dt}\left(\boldsymbol{\mathcal{P}}_{\boldsymbol{\mathcal{S}}}\left(\boldsymbol{X}_L(\epsilon^{-\alpha}t)\right)\boldsymbol{X}_L(\epsilon^{-\alpha}t)\right)$ yields a term proportional to the Dirac delta function in the limit $\epsilon \to 0$.

Define the n-dimensional vector of functions

$$\boldsymbol{\delta}_{L,\epsilon}(t) = \begin{cases} \dfrac{d}{dt}\left(\boldsymbol{\mathcal{P}}_{\boldsymbol{\mathcal{S}}}\left(\boldsymbol{X}_L\left(\epsilon^{-\alpha}t\right)\right)\boldsymbol{X}_L\left(\epsilon^{-\alpha}t\right)\right), & t \geq 0, \\ \dfrac{d}{d\tilde{t}}\left(\boldsymbol{\mathcal{P}}_{\boldsymbol{\mathcal{S}}}\left(\boldsymbol{X}_L\left(\epsilon^{-\alpha}\tilde{t}\right)\right)\boldsymbol{X}_L\left(\epsilon^{-\alpha}\tilde{t}\right)\right)\Big|_{\tilde{t}=-t}, & t < 0. \end{cases} \tag{4.351}$$

The function $\boldsymbol{\delta}_{L,\epsilon}(t)$ is continuous for $t = 0$ in every component. It can also be expressed as

$$\begin{aligned}\boldsymbol{\delta}_{L,\epsilon}(t) = {} & \epsilon^{-\alpha}\left(\nabla\boldsymbol{\mathcal{P}}_{\boldsymbol{\mathcal{S}}}\left(\boldsymbol{X}_L\left(\epsilon^{-\alpha}|t|\right)\right)\boldsymbol{X}'_L\left(\epsilon^{-\alpha}|t|\right)\right)\boldsymbol{X}_L\left(\epsilon^{-\alpha}|t|\right) \\ & + \epsilon^{-\alpha}\boldsymbol{\mathcal{P}}_{\boldsymbol{\mathcal{S}}}\left(\boldsymbol{X}_L\left(\epsilon^{-\alpha}|t|\right)\right)\boldsymbol{X}'_L\left(\epsilon^{-\alpha}|t|\right).\end{aligned} \tag{4.352}$$

First, evaluating $\boldsymbol{\delta}_{L,\epsilon}(t)$ at $t = 0$ yields

$$\boldsymbol{\delta}_{L,\epsilon}(0) = \epsilon^{-\alpha}\left(\left(\nabla\boldsymbol{\mathcal{P}}_{\boldsymbol{\mathcal{S}}}(\boldsymbol{x}_0)\boldsymbol{X}'_L(0)\right)\boldsymbol{x}_0 + \boldsymbol{\mathcal{P}}_{\boldsymbol{\mathcal{S}}}(\boldsymbol{x}_0)\boldsymbol{X}'_L(0)\right), \tag{4.353}$$

and because $\boldsymbol{X}'_L(0)$ is finite and does not depend on ϵ, this expression clearly diverges in the limit $\epsilon \to 0$,

$$\lim_{\epsilon\to 0}\boldsymbol{\delta}_{L,\epsilon}(0) = \infty\left(\left(\nabla\boldsymbol{\mathcal{P}}_{\boldsymbol{\mathcal{S}}}(\boldsymbol{x}_0)\boldsymbol{X}'_L(0)\right)\boldsymbol{x}_0 + \boldsymbol{\mathcal{P}}_{\boldsymbol{\mathcal{S}}}(\boldsymbol{x}_0)\boldsymbol{X}'_L(0)\right). \tag{4.354}$$

Second, for $|t| > 0$, $\lim_{\epsilon\to 0}\boldsymbol{\delta}_{L,\epsilon}(t)$ behaves as

$$\lim_{\epsilon\to 0}\boldsymbol{\delta}_{L,\epsilon}(t) = \mathbf{0},\ t \neq 0, \tag{4.355}$$

because $\boldsymbol{X}'_L\left(\epsilon^{-\alpha}|t|\right)$ appears in both terms of Eq. (4.352) and behaves as (see also Eq. (4.325))

$$\lim_{\epsilon\to 0}\boldsymbol{X}'_L\left(\epsilon^{-\alpha}|t|\right) = 0,\ t \neq 0. \tag{4.356}$$

Third, the integral of $\boldsymbol{\delta}_{L,\epsilon}(t)$ over time t must be determined. The integral can be split up in two integrals,

$$\begin{aligned}\int_{-\infty}^{\infty} d\tilde{t}\boldsymbol{\delta}_{L,\epsilon}\left(\tilde{t}\right) &= \int_{-\infty}^{0} d\tilde{t}\boldsymbol{\delta}_{L,\epsilon}\left(\tilde{t}\right) + \int_{0}^{\infty} d\tilde{t}\boldsymbol{\delta}_{L,\epsilon}\left(\tilde{t}\right) \\ &= \epsilon^{-\alpha}\int_{-\infty}^{0} d\tilde{t}\left(\nabla\boldsymbol{\mathcal{P}}_{\boldsymbol{\mathcal{S}}}\left(\boldsymbol{X}_L\left(-\epsilon^{-\alpha}\tilde{t}\right)\right)\boldsymbol{X}'_L\left(-\epsilon^{-\alpha}\tilde{t}\right)\right)\boldsymbol{X}_L\left(-\epsilon^{-\alpha}\tilde{t}\right) \\ &\quad + \epsilon^{-\alpha}\int_{-\infty}^{0} d\tilde{t}\boldsymbol{\mathcal{P}}_{\boldsymbol{\mathcal{S}}}\left(\boldsymbol{X}_L\left(-\epsilon^{-\alpha}\tilde{t}\right)\right)\boldsymbol{X}'_L\left(-\epsilon^{-\alpha}\tilde{t}\right) \\ &\quad + \epsilon^{-\alpha}\int_{0}^{\infty} d\tilde{t}\left(\nabla\boldsymbol{\mathcal{P}}_{\boldsymbol{\mathcal{S}}}\left(\boldsymbol{X}_L\left(\epsilon^{-\alpha}\tilde{t}\right)\right)\boldsymbol{X}'_L\left(\epsilon^{-\alpha}\tilde{t}\right)\right)\boldsymbol{X}_L\left(\epsilon^{-\alpha}\tilde{t}\right) \\ &\quad + \epsilon^{-\alpha}\int_{0}^{\infty} d\tilde{t}\boldsymbol{\mathcal{P}}_{\boldsymbol{\mathcal{S}}}\left(\boldsymbol{X}_L\left(\epsilon^{-\alpha}\tilde{t}\right)\right)\boldsymbol{X}'_L\left(\epsilon^{-\alpha}\tilde{t}\right).\end{aligned} \tag{4.357}$$

Substituting $\tau = -\epsilon^{-\alpha}\tilde{t}$ in the first and $\tau = \epsilon^{-\alpha}\tilde{t}$ in the second integral yields

$$\begin{aligned}\int_{-\infty}^{\infty} d\tilde{t}\boldsymbol{\delta}_{L,\epsilon}\left(\tilde{t}\right) &= 2\int_{0}^{\infty} d\tau\left(\left(\nabla\boldsymbol{\mathcal{P}}_{\boldsymbol{\mathcal{S}}}\left(\boldsymbol{X}_L\left(\tau\right)\right)\boldsymbol{X}'_L\left(\tau\right)\right)\boldsymbol{X}_L\left(\tau\right) + \boldsymbol{\mathcal{P}}_{\boldsymbol{\mathcal{S}}}\left(\boldsymbol{X}_L\left(\tau\right)\right)\boldsymbol{X}'_L\left(\tau\right)\right)\\ &= 2\left(\boldsymbol{\mathcal{P}}_{\boldsymbol{\mathcal{S}}}\left(\boldsymbol{x}_O\left(t_0\right)\right)\boldsymbol{x}_O\left(t_0\right) - \boldsymbol{\mathcal{P}}_{\boldsymbol{\mathcal{S}}}\left(\boldsymbol{x}_0\right)\boldsymbol{x}_0\right). \end{aligned} \tag{4.358}$$

Thus, we proved that

$$\lim_{\epsilon\to 0}\boldsymbol{\delta}_{L,\epsilon}\left(t\right) = 2\left(\boldsymbol{\mathcal{P}}_{\boldsymbol{\mathcal{S}}}\left(\boldsymbol{x}_O\left(t_0\right)\right)\boldsymbol{x}_O\left(t_0\right) - \boldsymbol{\mathcal{P}}_{\boldsymbol{\mathcal{S}}}\left(\boldsymbol{x}_0\right)\boldsymbol{x}_0\right)\delta\left(t\right). \tag{4.359}$$

Expressing the time derivative of $\boldsymbol{\mathcal{P}}_{\boldsymbol{\mathcal{S}}}\boldsymbol{X}_L$ as

$$\frac{d}{dt}\left(\boldsymbol{\mathcal{P}}_{\boldsymbol{\mathcal{S}}}\left(\boldsymbol{X}_L\left(\epsilon^{-\alpha}\left(t-t_0\right)\right)\right)\boldsymbol{X}_L\left(\epsilon^{-\alpha}\left(t-t_0\right)\right)\right) = \boldsymbol{\delta}_{L,\epsilon}\left(t-t_0\right),\, t\geq t_0, \tag{4.360}$$

finally gives

$$\begin{aligned}&\lim_{\epsilon\to 0}\frac{d}{dt}\left(\boldsymbol{\mathcal{P}}_{\boldsymbol{\mathcal{S}}}\left(\boldsymbol{X}_L\left(\epsilon^{-\alpha}\left(t-t_0\right)\right)\right)\boldsymbol{X}_L\left(\epsilon^{-\alpha}\left(t-t_0\right)\right)\right)\\ &= \lim_{\epsilon\to 0}\boldsymbol{\delta}_{L,\epsilon}\left(t-t_0\right)\\ &= 2\left(\boldsymbol{\mathcal{P}}_{\boldsymbol{\mathcal{S}}}\left(\boldsymbol{x}_O\left(t_0\right)\right)\boldsymbol{x}_O\left(t_0\right) - \boldsymbol{\mathcal{P}}_{\boldsymbol{\mathcal{S}}}\left(\boldsymbol{x}_0\right)\boldsymbol{x}_0\right)\delta\left(t-t_0\right). \end{aligned} \tag{4.361}$$

A similar discussion for the right inner equation yields the equivalent result

$$\begin{aligned}&\lim_{\epsilon\to 0}\frac{d}{dt}\left(\boldsymbol{\mathcal{P}}_{\boldsymbol{\mathcal{S}}}\left(\boldsymbol{X}_R\left(\epsilon^{-\alpha}\left(t_1-t\right)\right)\right)\boldsymbol{X}_L\left(\epsilon^{-\alpha}\left(t_1-t\right)\right)\right)\\ &= -2\left(\boldsymbol{\mathcal{P}}_{\boldsymbol{\mathcal{S}}}\left(\boldsymbol{x}_O\left(t_1\right)\right)\boldsymbol{x}_O\left(t_1\right) - \boldsymbol{\mathcal{P}}_{\boldsymbol{\mathcal{S}}}\left(\boldsymbol{x}_1\right)\boldsymbol{x}_1\right)\delta\left(t_1-t\right). \end{aligned} \tag{4.362}$$

Finally, the exact solution for the control signal for $\epsilon = 0$ reads as

$$\boldsymbol{u}\left(t\right) = \begin{cases} \boldsymbol{\mathcal{B}}^g_{\boldsymbol{\mathcal{S}}}\left(\boldsymbol{x}_0\right)\left(\dot{\boldsymbol{x}}_O\left(t_0\right) - \boldsymbol{R}\left(\boldsymbol{x}_0\right)\right) & \\ +2\boldsymbol{\mathcal{B}}^g_{\boldsymbol{\mathcal{S}}}\left(\boldsymbol{x}_0\right)\left(\boldsymbol{\mathcal{P}}_{\boldsymbol{\mathcal{S}}}\left(\boldsymbol{x}_O\left(t_0\right)\right)\boldsymbol{x}_O\left(t_0\right) - \boldsymbol{\mathcal{P}}_{\boldsymbol{\mathcal{S}}}\left(\boldsymbol{x}_0\right)\boldsymbol{x}_0\right)\delta\left(t-t_0\right), & t = t_0,\\ \boldsymbol{\mathcal{B}}^g_{\boldsymbol{\mathcal{S}}}\left(\boldsymbol{x}_O\left(t\right)\right)\left(\dot{\boldsymbol{x}}_O\left(t\right) - \boldsymbol{R}\left(\boldsymbol{x}_O\left(t\right)\right)\right), & t_0 < t < t_1,\\ \boldsymbol{\mathcal{B}}^g_{\boldsymbol{\mathcal{S}}}\left(\boldsymbol{x}_1\right)\left(\dot{\boldsymbol{x}}_O\left(t_1\right) - \boldsymbol{R}\left(\boldsymbol{x}_1\right)\right) & \\ -2\boldsymbol{\mathcal{B}}^g_{\boldsymbol{\mathcal{S}}}\left(\boldsymbol{x}_1\right)\left(\boldsymbol{\mathcal{P}}_{\boldsymbol{\mathcal{S}}}\left(\boldsymbol{x}_O\left(t_1\right)\right)\boldsymbol{x}_O\left(t_1\right) - \boldsymbol{\mathcal{P}}_{\boldsymbol{\mathcal{S}}}\left(\boldsymbol{x}_1\right)\boldsymbol{x}_1\right)\delta\left(t_1-t\right) & t = t_1. \end{cases} \tag{4.363}$$

In conclusion, the control diverges at the initial and terminal time, $t = t_0$ and $t = t_1$, respectively. The divergence is in form of a Dirac delta function. The delta kick has a direction in state space parallel to the jump of the discontinuous state components. The strength of the delta kick is twice the height of the jump. Inside the time domain, the control signal is continuous and finite and entirely given in terms of the outer solution $\boldsymbol{x}_O\left(t\right)$.

4.4.9. Linearizing assumption

The exact state solution for optimal trajectory tracking for $\epsilon = 0$ is given solely in terms of the outer equations (4.341) and (4.342). Although these equations are simpler than the full necessary optimality conditions, they are nevertheless nonlinear and cannot be solved easily. However, Eqs. (4.341) and (4.342) become linear if a linearizing assumption holds, and a solution in closed form can be given.

First, the matrix $\boldsymbol{\Omega}_{\boldsymbol{\mathcal{S}}}(\boldsymbol{x})$ is assumed to be constant,

$$\boldsymbol{\Omega}_{\boldsymbol{\mathcal{S}}}(\boldsymbol{x}) = \boldsymbol{\mathcal{B}}(\boldsymbol{x}) \left(\boldsymbol{\mathcal{B}}^T(\boldsymbol{x}) \boldsymbol{\mathcal{S}} \boldsymbol{\mathcal{B}}(\boldsymbol{x})\right)^{-1} \boldsymbol{\mathcal{B}}^T(\boldsymbol{x}) = \text{const.} = \boldsymbol{\Omega}_{\boldsymbol{\mathcal{S}}}. \tag{4.364}$$

Note that this assumption does neither imply a constant coupling matrix $\boldsymbol{\mathcal{B}}(\boldsymbol{x})$ nor a constant matrix $\boldsymbol{\Gamma}_{\boldsymbol{\mathcal{S}}}(\boldsymbol{x})$ defined in Eq. (4.212). Equation (4.364) implies constant projectors $\boldsymbol{\mathcal{P}}_{\boldsymbol{\mathcal{S}}}(\boldsymbol{x})$ and $\boldsymbol{\mathcal{Q}}_{\boldsymbol{\mathcal{S}}}(\boldsymbol{x})$,

$$\boldsymbol{\mathcal{P}}_{\boldsymbol{\mathcal{S}}}(\boldsymbol{x}) = \boldsymbol{\Omega}_{\boldsymbol{\mathcal{S}}}(\boldsymbol{x}) \boldsymbol{\mathcal{S}} = \text{const.} = \boldsymbol{\mathcal{P}}_{\boldsymbol{\mathcal{S}}}, \tag{4.365}$$

$$\boldsymbol{\mathcal{Q}}_{\boldsymbol{\mathcal{S}}}(\boldsymbol{x}) = \mathbf{1} - \boldsymbol{\mathcal{P}}_{\boldsymbol{\mathcal{S}}}(\boldsymbol{x}) = \text{const.} = \boldsymbol{\mathcal{Q}}_{\boldsymbol{\mathcal{S}}}, \tag{4.366}$$

and analogously constant transposed projectors $\boldsymbol{\mathcal{P}}_{\boldsymbol{\mathcal{S}}}^T(\boldsymbol{x})$ and $\boldsymbol{\mathcal{Q}}_{\boldsymbol{\mathcal{S}}}^T(\boldsymbol{x})$.

Second, the nonlinearity $\boldsymbol{R}(\boldsymbol{x})$ is assumed to have the following structure with respect to the control,

$$\boldsymbol{\mathcal{Q}}_{\boldsymbol{\mathcal{S}}} \boldsymbol{R}(\boldsymbol{x}) = \boldsymbol{\mathcal{Q}}_{\boldsymbol{\mathcal{S}}} \boldsymbol{\mathcal{A}} \boldsymbol{x} + \boldsymbol{\mathcal{Q}}_{\boldsymbol{\mathcal{S}}} \boldsymbol{b}, \tag{4.367}$$

with constant $n \times n$ matrix $\boldsymbol{\mathcal{A}}$ and constant n-component vector $\boldsymbol{b}$. Cast into a single sentence, assumption Eq. (4.367) states that the control signals act on the nonlinear equations of the system, and all other equations are linear. Note that the linearizing assumption Eqs. (4.364) and (4.367) differs from the linearizing assumption of Section 2.3 in that it involves the matrix of weights $\boldsymbol{\mathcal{S}}$.

The linearizing assumption implies that the part $\boldsymbol{\mathcal{Q}}_{\boldsymbol{\mathcal{S}}} \nabla \boldsymbol{R}$ of the Jacobi matrix is independent of the state and given by

$$\boldsymbol{\mathcal{Q}}_{\boldsymbol{\mathcal{S}}} \nabla \boldsymbol{R}(\boldsymbol{x}) = \boldsymbol{\mathcal{Q}}_{\boldsymbol{\mathcal{S}}} \boldsymbol{\mathcal{A}}. \tag{4.368}$$

Transposing yields

$$\nabla \boldsymbol{R}^T(\boldsymbol{x}) \boldsymbol{\mathcal{Q}}_{\boldsymbol{\mathcal{S}}}^T = \boldsymbol{\mathcal{A}}^T \boldsymbol{\mathcal{Q}}_{\boldsymbol{\mathcal{S}}}^T. \tag{4.369}$$

Furthermore, Eqs. (4.364) and (4.367) imply

$$\boldsymbol{\mathcal{Q}}_{\boldsymbol{\mathcal{S}}} \boldsymbol{\mathcal{W}}(\boldsymbol{x}, \boldsymbol{y}) = \mathbf{0}, \tag{4.370}$$

because of Eq. (4.238). Analogously, it follows that

$$\boldsymbol{\mathcal{Q}}_{\boldsymbol{\mathcal{S}}} \boldsymbol{\mathcal{U}}(\boldsymbol{x}) = \mathbf{0}, \qquad \boldsymbol{\mathcal{Q}}_{\boldsymbol{\mathcal{S}}} \boldsymbol{\mathcal{V}}(\boldsymbol{x}, \boldsymbol{y}) = \mathbf{0}, \tag{4.371}$$

for the matrices $\boldsymbol{\mathcal{U}}$ and $\boldsymbol{\mathcal{V}}$ defined in Eqs. (4.267) and (4.268), respectively.

Under the linearizing assumption, the outer equations (4.255), (4.257), and (4.258) become linear,

$$-\boldsymbol{\mathcal{Q}}_{\mathcal{S}}^{T}\dot{\boldsymbol{\lambda}}_{O}(t) = \boldsymbol{\mathcal{Q}}_{\mathcal{S}}^{T}\boldsymbol{\mathcal{A}}^{T}\boldsymbol{\mathcal{Q}}_{\mathcal{S}}^{T}\boldsymbol{\lambda}_{O}(t) + \boldsymbol{\mathcal{Q}}_{\mathcal{S}}^{T}\boldsymbol{\mathcal{S}}\boldsymbol{\mathcal{Q}}_{\mathcal{S}}\left(\boldsymbol{x}_{O}(t) - \boldsymbol{x}_{d}(t)\right), \tag{4.372}$$

$$\boldsymbol{\mathcal{P}}_{\mathcal{S}}^{T}\boldsymbol{\lambda}_{O}(t) = \boldsymbol{0}, \tag{4.373}$$

$$\boldsymbol{\mathcal{P}}_{\mathcal{S}}\boldsymbol{x}_{O}(t) = \boldsymbol{\mathcal{P}}_{\mathcal{S}}\boldsymbol{x}_{d}(t) - \boldsymbol{\Omega}_{\mathcal{S}}\boldsymbol{\mathcal{A}}^{T}\boldsymbol{\mathcal{Q}}_{\mathcal{S}}^{T}\boldsymbol{\lambda}_{O}(t), \tag{4.374}$$

$$\boldsymbol{\mathcal{Q}}_{\mathcal{S}}\dot{\boldsymbol{x}}_{O}(t) = \boldsymbol{\mathcal{Q}}_{\mathcal{S}}\boldsymbol{\mathcal{A}}\boldsymbol{x}_{O}(t). \tag{4.375}$$

The system of Eqs. (4.372)-(4.375) is linear and can be solved. Using Eq. (4.374), $\boldsymbol{\mathcal{P}}_{\mathcal{S}}\boldsymbol{x}_{O}(t)$ can be eliminated from Eq. (4.375). This yields a system of $2(n-p)$ inhomogeneous first order ODEs for $\boldsymbol{\mathcal{Q}}_{\mathcal{S}}^{T}\boldsymbol{\lambda}_{O}(t)$ and $\boldsymbol{\mathcal{Q}}_{\mathcal{S}}\boldsymbol{x}_{O}(t)$,

$$\begin{pmatrix} \boldsymbol{\mathcal{Q}}_{\mathcal{S}}^{T}\dot{\boldsymbol{\lambda}}_{O}(t) \\ \boldsymbol{\mathcal{Q}}_{\mathcal{S}}\dot{\boldsymbol{x}}_{O}(t) \end{pmatrix} = \begin{pmatrix} -\boldsymbol{\mathcal{Q}}_{\mathcal{S}}^{T}\boldsymbol{\mathcal{A}}^{T}\boldsymbol{\mathcal{Q}}_{\mathcal{S}}^{T} & -\boldsymbol{\mathcal{Q}}_{\mathcal{S}}^{T}\boldsymbol{\mathcal{S}}\boldsymbol{\mathcal{Q}}_{\mathcal{S}} \\ -\boldsymbol{\mathcal{Q}}_{\mathcal{S}}\boldsymbol{\mathcal{A}}\boldsymbol{\Omega}_{\mathcal{S}}\boldsymbol{\mathcal{A}}^{T}\boldsymbol{\mathcal{Q}}_{\mathcal{S}}^{T} & \boldsymbol{\mathcal{Q}}_{\mathcal{S}}\boldsymbol{\mathcal{A}}\boldsymbol{\mathcal{Q}}_{\mathcal{S}} \end{pmatrix} \begin{pmatrix} \boldsymbol{\mathcal{Q}}_{\mathcal{S}}^{T}\boldsymbol{\lambda}_{O}(t) \\ \boldsymbol{\mathcal{Q}}_{\mathcal{S}}\boldsymbol{x}_{O}(t) \end{pmatrix} + \begin{pmatrix} \boldsymbol{\mathcal{Q}}_{\mathcal{S}}^{T}\boldsymbol{\mathcal{S}}\boldsymbol{\mathcal{Q}}_{\mathcal{S}}\boldsymbol{x}_{d}(t) \\ \boldsymbol{\mathcal{Q}}_{\mathcal{S}}\boldsymbol{\mathcal{A}}\boldsymbol{\mathcal{P}}_{\mathcal{S}}\boldsymbol{x}_{d}(t) + \boldsymbol{\mathcal{Q}}_{\mathcal{S}}\boldsymbol{b} \end{pmatrix}. \tag{4.376}$$

Equation (4.376) has to be solved with the initial and terminal conditions

$$\boldsymbol{\mathcal{Q}}_{\mathcal{S}}\boldsymbol{x}(t_0) = \boldsymbol{\mathcal{Q}}_{\mathcal{S}}\boldsymbol{x}_0, \qquad \boldsymbol{\mathcal{Q}}_{\mathcal{S}}\boldsymbol{x}(t_1) = \boldsymbol{\mathcal{Q}}_{\mathcal{S}}\boldsymbol{x}_1. \tag{4.377}$$

The solution to Eq. (4.376) can be expressed in closed form in terms of the state transition matrix $\boldsymbol{\Phi}(t, t_0)$,

$$\begin{pmatrix} \boldsymbol{\mathcal{Q}}_{\mathcal{S}}^{T}\boldsymbol{\lambda}_{O}(t) \\ \boldsymbol{\mathcal{Q}}_{\mathcal{S}}\boldsymbol{x}_{O}(t) \end{pmatrix} = \boldsymbol{\Phi}(t, t_0) \begin{pmatrix} \boldsymbol{\mathcal{Q}}_{\mathcal{S}}^{T}\boldsymbol{\lambda}_{\text{init}} \\ \boldsymbol{\mathcal{Q}}_{\mathcal{S}}\boldsymbol{x}_0 \end{pmatrix} + \int_{t_0}^{t} d\tau \boldsymbol{\Phi}(t, \tau) \begin{pmatrix} \boldsymbol{\mathcal{Q}}_{\mathcal{S}}^{T}\boldsymbol{\mathcal{S}}\boldsymbol{\mathcal{Q}}_{\mathcal{S}}\boldsymbol{x}_{d}(\tau) \\ \boldsymbol{\mathcal{Q}}_{\mathcal{S}}\boldsymbol{\mathcal{A}}\boldsymbol{\mathcal{P}}_{\mathcal{S}}\boldsymbol{x}_{d}(\tau) + \boldsymbol{\mathcal{Q}}_{\mathcal{S}}\boldsymbol{b} \end{pmatrix}, \tag{4.378}$$

see Appendix A.1. The term $\boldsymbol{\mathcal{Q}}_{\mathcal{S}}^{T}\boldsymbol{\lambda}_{\text{init}}$ must be determined by the terminal condition $\boldsymbol{\mathcal{Q}}_{\mathcal{S}}\boldsymbol{x}(t_1) = \boldsymbol{\mathcal{Q}}_{\mathcal{S}}\boldsymbol{x}_1$. Because the state matrix of Eq. (4.376) is constant in time, the state transition matrix is given by the matrix exponential and can be formally written as

$$\boldsymbol{\Phi}(t, t_0) = \exp\left(\begin{pmatrix} -\boldsymbol{\mathcal{Q}}_{\mathcal{S}}^{T}\boldsymbol{\mathcal{A}}^{T}\boldsymbol{\mathcal{Q}}_{\mathcal{S}}^{T} & -\boldsymbol{\mathcal{Q}}_{\mathcal{S}}^{T}\boldsymbol{\mathcal{S}}\boldsymbol{\mathcal{Q}}_{\mathcal{S}} \\ -\boldsymbol{\mathcal{Q}}_{\mathcal{S}}\boldsymbol{\mathcal{A}}\boldsymbol{\Omega}_{\mathcal{S}}\boldsymbol{\mathcal{A}}^{T}\boldsymbol{\mathcal{Q}}_{\mathcal{S}}^{T} & \boldsymbol{\mathcal{Q}}_{\mathcal{S}}\boldsymbol{\mathcal{A}}\boldsymbol{\mathcal{Q}}_{\mathcal{S}} \end{pmatrix}(t - t_0)\right). \tag{4.379}$$

We emphasize that the linearizing assumption only leads to linear outer equations. In general, the inner equations are nonlinear even if the linearizing assumption holds. This is demonstrated for the Case 2.2 of inner equations, see Sections 4.4.3.3 and 4.4.4.3. Case 2.2 corresponds to the left and right inner equations of the two-dimensional dynamical system from Section 4.1. The left inner equations (4.285) and (4.287) become

$$\boldsymbol{\mathcal{Q}}_{\mathcal{S}}^{T}\boldsymbol{\Lambda}'_{L}(\tau_L) = \boldsymbol{0}, \qquad \boldsymbol{\mathcal{Q}}_{\mathcal{S}}\boldsymbol{X}'_{L}(\tau_L) = \boldsymbol{0}. \tag{4.380}$$

The initial conditions Eqs. (4.310) and (4.313) lead to

$$\boldsymbol{\mathcal{Q}}_{\boldsymbol{\mathcal{S}}}^{T}\boldsymbol{\Lambda}_{L}\left(\tau_{L}\right)=\boldsymbol{\mathcal{Q}}_{\boldsymbol{\mathcal{S}}}^{T}\boldsymbol{\lambda}_{O}\left(t_{0}\right),\qquad\qquad\boldsymbol{\mathcal{Q}}_{\boldsymbol{\mathcal{S}}}\boldsymbol{X}_{L}\left(\tau_{L}\right)=\boldsymbol{\mathcal{Q}}_{\boldsymbol{\mathcal{S}}}\boldsymbol{x}_{0},\tag{4.381}$$

and the remaining left inner equation (4.286) becomes

$$\begin{aligned}\frac{\partial}{\partial\tau_{L}}\left(\boldsymbol{\Gamma}_{\boldsymbol{\mathcal{S}}}\left(\boldsymbol{X}_{L}\left(\tau_{L}\right)\right)\boldsymbol{X}_{L}'\left(\tau_{L}\right)\right)&=-\boldsymbol{\mathcal{P}}_{\boldsymbol{\mathcal{S}}}^{T}\boldsymbol{\mathcal{V}}^{T}\left(\boldsymbol{X}_{L}\left(\tau_{L}\right),\boldsymbol{X}_{L}'\left(\tau_{L}\right)\right)\boldsymbol{\Gamma}_{\boldsymbol{\mathcal{S}}}\left(\boldsymbol{X}_{L}\left(\tau_{L}\right)\right)\boldsymbol{X}_{L}'\left(\tau_{L}\right)\\&\quad+\boldsymbol{\mathcal{P}}_{\boldsymbol{\mathcal{S}}}^{T}\boldsymbol{\mathcal{A}}^{T}\boldsymbol{\mathcal{Q}}_{\boldsymbol{\mathcal{S}}}^{T}\boldsymbol{\lambda}_{O}\left(t_{0}\right)+\boldsymbol{\mathcal{P}}_{\boldsymbol{\mathcal{S}}}^{T}\boldsymbol{\mathcal{S}}\boldsymbol{\mathcal{P}}_{\boldsymbol{\mathcal{S}}}\left(\boldsymbol{X}_{L}\left(\tau_{L}\right)-\boldsymbol{x}_{d}\left(t_{0}\right)\right).\end{aligned}\tag{4.382}$$

In general, the matrices $\boldsymbol{\Gamma}_{\boldsymbol{\mathcal{S}}}$ and $\boldsymbol{\mathcal{V}}$ depend nonlinearly on the state. In both cases, the nonlinearity originates from the coupling matrix $\boldsymbol{\mathcal{B}}\left(\boldsymbol{x}\right)$, see Eqs. (4.212) and (4.268), and no trace is left by the nonlinearity $\boldsymbol{R}\left(\boldsymbol{x}\right)$. Equation (4.382) has to be solved with the initial condition

$$\boldsymbol{\mathcal{P}}_{\boldsymbol{\mathcal{S}}}\boldsymbol{X}_{L}\left(0\right)=\boldsymbol{\mathcal{P}}_{\boldsymbol{\mathcal{S}}}\boldsymbol{x}_{0}.\tag{4.383}$$

The terminal condition, Eq. (4.324), becomes

$$\lim_{\tau_{L}\rightarrow\infty}\boldsymbol{\mathcal{P}}_{\boldsymbol{\mathcal{S}}}\boldsymbol{X}_{L}\left(\tau_{L}\right)=\boldsymbol{\mathcal{P}}_{\boldsymbol{\mathcal{S}}}\boldsymbol{x}_{d}\left(t_{0}\right)-\boldsymbol{\Omega}_{\boldsymbol{\mathcal{S}}}\boldsymbol{\mathcal{A}}^{T}\boldsymbol{\mathcal{Q}}_{\boldsymbol{\mathcal{S}}}^{T}\boldsymbol{\lambda}_{O}\left(t_{0}\right).\tag{4.384}$$

The existence of the limit Eq. (4.384) implies

$$\lim_{\tau_{L}\rightarrow\infty}\boldsymbol{\mathcal{P}}_{\boldsymbol{\mathcal{S}}}\boldsymbol{X}_{L}'\left(\tau_{L}\right)=\boldsymbol{0}.\tag{4.385}$$

Comparing this limit with the limit $\tau_{L}\rightarrow\infty$ of Eq. (4.382), multiplied by $\boldsymbol{\Omega}_{\boldsymbol{\mathcal{S}}}$ from the left, yields indeed Eq. (4.384).

The right inner equations (4.303) and (4.305) become

$$\boldsymbol{\mathcal{Q}}_{\boldsymbol{\mathcal{S}}}^{T}\boldsymbol{\Lambda}_{R}'\left(\tau_{R}\right)=\boldsymbol{0},\qquad\qquad\boldsymbol{\mathcal{Q}}_{\boldsymbol{\mathcal{S}}}\boldsymbol{X}_{R}'\left(\tau_{R}\right)=\boldsymbol{0}.\tag{4.386}$$

The initial conditions Eqs. (4.311) and (4.316) lead to

$$\boldsymbol{\mathcal{Q}}_{\boldsymbol{\mathcal{S}}}^{T}\boldsymbol{\Lambda}_{R}\left(\tau_{R}\right)=\boldsymbol{\mathcal{Q}}_{\boldsymbol{\mathcal{S}}}^{T}\boldsymbol{\lambda}_{O}\left(t_{1}\right),\qquad\qquad\boldsymbol{\mathcal{Q}}_{\boldsymbol{\mathcal{S}}}\boldsymbol{X}_{R}\left(\tau_{R}\right)=\boldsymbol{\mathcal{Q}}_{\boldsymbol{\mathcal{S}}}\boldsymbol{x}_{1}.\tag{4.387}$$

The remaining right inner equation (4.304) is

$$\begin{aligned}\frac{\partial}{\partial\tau_{R}}\left(\boldsymbol{\Gamma}_{\boldsymbol{\mathcal{S}}}\left(\boldsymbol{X}_{R}\left(\tau_{R}\right)\right)\boldsymbol{X}_{R}'\left(\tau_{R}\right)\right)&=-\boldsymbol{\mathcal{P}}_{\boldsymbol{\mathcal{S}}}^{T}\boldsymbol{\mathcal{V}}^{T}\left(\boldsymbol{X}_{R}\left(\tau_{R}\right),\boldsymbol{X}_{R}'\left(\tau_{R}\right)\right)\boldsymbol{\Gamma}_{\boldsymbol{\mathcal{S}}}\left(\boldsymbol{X}_{R}\left(\tau_{R}\right)\right)\boldsymbol{X}_{R}'\left(\tau_{R}\right)\\&\quad+\boldsymbol{\mathcal{P}}_{\boldsymbol{\mathcal{S}}}^{T}\boldsymbol{\mathcal{A}}^{T}\boldsymbol{\mathcal{Q}}_{\boldsymbol{\mathcal{S}}}^{T}\boldsymbol{\lambda}_{O}\left(t_{1}\right)+\boldsymbol{\mathcal{P}}_{\boldsymbol{\mathcal{S}}}^{T}\boldsymbol{\mathcal{S}}\boldsymbol{\mathcal{P}}_{\boldsymbol{\mathcal{S}}}\left(\boldsymbol{X}_{R}\left(\tau_{R}\right)-\boldsymbol{x}_{d}\left(t_{1}\right)\right),\end{aligned}\tag{4.388}$$

which is to be solved together with the initial condition

$$\boldsymbol{\mathcal{P}}_{\boldsymbol{\mathcal{S}}}\boldsymbol{X}_{R}\left(0\right)=\boldsymbol{\mathcal{P}}_{\boldsymbol{\mathcal{S}}}\boldsymbol{x}_{1}.\tag{4.389}$$

Similar as for the left side, the terminal condition, Eq. (4.327),

$$\lim_{\tau_R \to \infty} \boldsymbol{\mathcal{P}}_{\boldsymbol{\mathcal{S}}} \boldsymbol{X}_R(\tau_R) = \boldsymbol{\mathcal{P}}_{\boldsymbol{\mathcal{S}}} \boldsymbol{x}_d(t_1) - \boldsymbol{\Omega}_{\boldsymbol{\mathcal{S}}} \boldsymbol{\mathcal{A}}^T \boldsymbol{\mathcal{Q}}_{\boldsymbol{\mathcal{S}}}^T \boldsymbol{\lambda}_O(t_1) \tag{4.390}$$

is already satisfied because Eq. (4.390) implies

$$\lim_{\tau_R \to \infty} \boldsymbol{\mathcal{P}}_{\boldsymbol{\mathcal{S}}} \boldsymbol{X}'_R(\tau_R) = \boldsymbol{0}. \tag{4.391}$$

The solution for the control signal is given by

$$\boldsymbol{u}(t) = \begin{cases} \boldsymbol{\mathcal{B}}_{\boldsymbol{\mathcal{S}}}^g(\boldsymbol{x}_0)\left(\dot{\boldsymbol{x}}_O(t_0) - \boldsymbol{R}(\boldsymbol{x}_0) + 2\left(\boldsymbol{x}_O(t_0) - \boldsymbol{x}_0\right)\delta(t - t_0)\right), & t = t_0, \\ \boldsymbol{\mathcal{B}}_{\boldsymbol{\mathcal{S}}}^g(\boldsymbol{x}_O(t))\left(\dot{\boldsymbol{x}}_O(t) - \boldsymbol{R}(\boldsymbol{x}_O(t))\right), & t_0 < t < t_1, \\ \boldsymbol{\mathcal{B}}_{\boldsymbol{\mathcal{S}}}^g(\boldsymbol{x}_1)\left(\dot{\boldsymbol{x}}_O(t_1) - \boldsymbol{R}(\boldsymbol{x}_1) - 2\left(\boldsymbol{x}_O(t_1) - \boldsymbol{x}_1\right)\delta(t_1 - t)\right) & t = t_1. \end{cases} \tag{4.392}$$

In conclusion, for $\epsilon = 0$ and valid linearizing assumption, the exact state solution is given by the linear outer equations (4.376) accompanied by jumps at the time domain boundaries. Thus, the analytical *form* of the exact solution is solely determined by linear equations. However, the *existence* of the exact solution relies not only on the existence of outer solutions to Eq. (4.376), but also on the existence of solutions to the generally nonlinear inner equations (4.382) and (4.388). Only the existence of inner solutions guarantees the existence of jumps connecting the initial and terminal conditions with the outer solution. It is in this sense that we are able to speak about an underlying linear structure of nonlinear optimal trajectory tracking.

4.4.10. Discussion

Analytical approximations for optimal trajectory tracking of nonlinear affine dynamical systems are developed in this section. In contrast to Chapter 2, which discusses only exactly realizable trajectories, the results given here are valid for arbitrary desired trajectories. The general structure of the solution for small regularization parameter $0 \leq \epsilon \ll 1$ is unveiled. The n state components $\boldsymbol{x}(t)$ are separated by the two complementary projectors $\boldsymbol{\mathcal{P}}_{\boldsymbol{\mathcal{S}}}(\boldsymbol{x})$ and $\boldsymbol{\mathcal{Q}}_{\boldsymbol{\mathcal{S}}}(\boldsymbol{x})$, while the n co-state components $\boldsymbol{\lambda}(t)$ are separated by the transposed projectors $\boldsymbol{\mathcal{P}}_{\boldsymbol{\mathcal{S}}}^T(\boldsymbol{x})$ and $\boldsymbol{\mathcal{Q}}_{\boldsymbol{\mathcal{S}}}^T(\boldsymbol{x})$.

For all $\epsilon > 0$, the dynamics of an optimal control system takes place in the combined state space of dimension $2n$ of state $\boldsymbol{x}$ and co-state $\boldsymbol{\lambda}$ and is governed by $2n$ first order ODEs. The exact solution for $\epsilon = 0$ is governed by $2(n-p)$ first order ODEs for the state components $\boldsymbol{\mathcal{Q}}_{\boldsymbol{\mathcal{S}}}(\boldsymbol{x}(t))\,\boldsymbol{x}(t)$ and the co-state components $\boldsymbol{\mathcal{Q}}_{\boldsymbol{\mathcal{S}}}^T(\boldsymbol{x}(t))\,\boldsymbol{\lambda}(t)$, respectively. These equations are called the outer equations and given by (4.341) and (4.342), respectively. The $2p$ state components $\boldsymbol{\mathcal{P}}_{\boldsymbol{\mathcal{S}}}(\boldsymbol{x}(t))\,\boldsymbol{x}(t)$ and co-state components $\boldsymbol{\mathcal{P}}_{\boldsymbol{\mathcal{S}}}^T(\boldsymbol{x}(t))\,\boldsymbol{\lambda}(t)$ are given by algebraic equations. The part $\boldsymbol{\mathcal{P}}_{\boldsymbol{\mathcal{S}}}^T(\boldsymbol{x}(t))\,\boldsymbol{\lambda}(t) = \boldsymbol{0}$ vanishes for all times, while the part $\boldsymbol{\mathcal{P}}_{\boldsymbol{\mathcal{S}}}(\boldsymbol{x}(t))\,\boldsymbol{x}(t)$ is given by the algebraic equation (4.338) inside the time domain, $t_0 < t < t_1$. Th $2p$ algebraic

equations restrict the dynamics to a hypersurface of dimension $2\,(n-p)$ embedded in the extended phase space of dimension $2n$. For all times except for the beginning, $t = t_0$, and end of the time interval, $t = t_1$, the system is evolving on the so-called *singular surface* (Bryson and Ho, 1975). Note that the singular surface is time dependent if the desired trajectory $\boldsymbol{x}_d\,(t)$ is time dependent.

The outer equations (4.341) and (4.342) are $2\,(n-p)$ first order ODEs which allow for $2\,(n-p)$ initial conditions. This is not enough to accommodate all $2n$ initial and terminal conditions given by Eq. (4.201). For $\epsilon = 0$, this results in instantaneous and discontinuous transitions, or jumps, at the time domain boundaries. At $t = t_0$, a jump from the initial condition $\boldsymbol{x}_0$ onto the singular surface occurs. Similarly, at $t = t_1$, a jump from the singular surface onto the terminal condition $\boldsymbol{x}_1$ takes place. These jumps manifest as discontinuities in the $2p$ state components $\boldsymbol{\mathcal{P}}_{\boldsymbol{\mathcal{S}}}\,(\boldsymbol{x}\,(t))\,\boldsymbol{x}\,(t)$. The heights and directions of the jumps, see Eqs. (4.346) and (4.347), are given by differences between the initial and terminal conditions and the initial and terminal values of the outer solutions, respectively. The jumps are mediated by control impulses in form of Dirac delta functions located at the beginning and the end of the time interval. The direction of a delta kick, given by the coefficient of the Dirac delta function, is parallel to the direction of the jump occurring at the same instant. The strength of the kick is twice the height of the jump. Intuitively, the reason is that the delta kicks are located right at the time domain boundaries such that only half of the kicks contribute to the time evolution.

Section 4.4.7 demonstrates that the exact solution for $\epsilon = 0$ is entirely expressed in terms of the outer solutions given by Eqs. (4.341), (4.342), and (4.338). No trace remains of the inner solutions except the mere existence and location of the jumps at the time domain boundaries. Section 4.4.9 draws the final conclusion and states sufficient conditions for the linearity of the outer equations. Because of their linearity, a formal closed form solution valid for arbitrary desired trajectories $\boldsymbol{x}_d\,(t)$ can be given. This establishes an underlying linear structure of nonlinear unregularized optimal trajectory tracking for affine control systems satisfying the linearizing assumption. This finding constitutes the major result of this thesis.

While the form of the exact state trajectory for $\epsilon = 0$ is entirely given by the outer equations, its existence relies on the existence of solutions to the inner equations for Case 1 and Case 2.2. The existence of inner solutions for appropriate initial, terminal, and matching conditions ensures the existence of jumps connecting initial and terminal conditions with the singular surface. If inner solutions do not exist for Case 1 and Case 2.2, different or additional scaling regimes may exist which ensure the existence of jumps. The linearizing assumption eliminates all nonlinear terms originating from the nonlinearity $\boldsymbol{R}\,(\boldsymbol{x})$ in the outer equations and the inner equations. However, the inner equations are generally nonlinear, with nonlinear terms originating from a state-dependent coupling matrix $\boldsymbol{\mathcal{B}}\,(\boldsymbol{x})$. For a constant coupling matrix, the inner equations are linear as well. In conclusion, if $\epsilon = 0$ and the linearizing assumption holds, the form of the exact state and co-state trajectory is given by linear ODEs, but their existence relies on additional, generally nonlinear

ODEs.

Having obtained analytical results for optimal open loop control, it is in principle possible to extend this result to continuous time and continuous time-delayed feedback control. The computations proceed along the same lines as in Section 4.3 by promoting the initial state to a functional of the controlled state. For the perturbation expansion in this section, sharp terminal conditions $\boldsymbol{x}(t_1) = \boldsymbol{x}_1$ were assumed. Analytically, this is the simplest choice, but requires the system to be controllable. An extension to more general terminal conditions is desirable but not straightforward.

The projectors $\boldsymbol{\mathcal{P}}_{\boldsymbol{\mathcal{S}}}$ and $\boldsymbol{\mathcal{Q}}_{\boldsymbol{\mathcal{S}}}$ play an essential role for the solution. Both projectors are derived with the help of the generalized Legendre-Clebsch condition for singular optimal control in Section 3.4.2. While $\boldsymbol{\mathcal{P}}_{\boldsymbol{\mathcal{S}}}$ depends on the matrix of weighting coefficients $\boldsymbol{\mathcal{S}}$, the projector $\boldsymbol{\mathcal{P}}$ defined in Chapter 2 is independent of $\boldsymbol{\mathcal{S}}$ and $\boldsymbol{\mathcal{P}}_{\boldsymbol{\mathcal{S}}}$ reduces to $\boldsymbol{\mathcal{P}}$ for $\boldsymbol{\mathcal{S}} = \mathbf{1}$. The necessity to use $\boldsymbol{\mathcal{P}}_{\boldsymbol{\mathcal{S}}}$ instead of $\boldsymbol{\mathcal{P}}$ becomes obvious in the derivation of Eq. (4.258), which cannot be obtained with projector $\boldsymbol{\mathcal{P}}$. For exactly realizable desired trajectories $\boldsymbol{x}_d(t)$, it is irrelevant which projector is used. The control and controlled state obtained with $\boldsymbol{\mathcal{P}}_{\boldsymbol{\mathcal{S}}}$ are independent of $\boldsymbol{\mathcal{S}}$ and identical to results obtained with $\boldsymbol{\mathcal{P}}$, see Section 3.4.2 for a proof. In contrast, optimal trajectory tracking for arbitrary desired trajectories $\boldsymbol{x}_d(t)$ yields a control signal and controlled state trajectory which depends explicitly on $\boldsymbol{\mathcal{S}}$. Analogously, the linearizing assumption introduced in Section 4.4.9 depends on $\boldsymbol{\mathcal{S}}$ and is different from the linearizing assumption in Section 2.3. However, the matrices $\boldsymbol{\mathcal{Q}}_{\boldsymbol{\mathcal{S}}}$ and $\boldsymbol{\mathcal{Q}}$ are similar and have identical diagonal representations. This suggests that if a linearizing assumption holds in terms of $\boldsymbol{\mathcal{Q}}_{\boldsymbol{\mathcal{S}}}$, it also holds in terms of $\boldsymbol{\mathcal{Q}}$, and vice versa. A rigorous proof of this conjecture is desirable.

4.5. Conclusions

4.5.1. Analytical results for $\epsilon \to 0$

Analytical approximations for optimal trajectory tracking in nonlinear affine control systems were derived in this chapter. The regularization parameter ϵ is used as the small parameter for a perturbation expansion, and the solutions become exact for $\epsilon = 0$. As discussed in Section 3.1.4, the case $\epsilon = 0$ can be seen as the limit of realizability of a certain desired trajectory $\boldsymbol{x}_d(t)$. No other control, be it open or closed loop control, can enforce a state trajectory $\boldsymbol{x}(t)$ with a smaller distance to the desired state trajectory $\boldsymbol{x}_d(t)$. Importantly, the regularization parameter originates solely from the formulation of the control problem. The system dynamics is exactly taken into account. The analytical approximations do neither require any simplifying assumptions about the strength of nonlinearities, as e.g. weak nonlinearities, nor about the separation of time scales between different state components,

or similar. To solve the equations derived by the perturbative treatment in closed form, however, the nonlinearity $\boldsymbol{R}(\boldsymbol{x})$ has to have a simple structure with respect to the coupling matrix $\boldsymbol{\mathcal{B}}(\boldsymbol{x})$. This structure is defined in abstract notation valid for a general affine control system in Eqs. (4.364) and (4.367) and called the linearizing assumption. Cast in words, this assumption becomes "the control signals act on the state components governed by nonlinear equations, and all other components are governed by linear equations". The linearizing assumption results in linear equations for unregularized nonlinear optimal trajectory tracking. Only due to this linearity it is possible to derive solutions in closed form valid for arbitrary desired trajectories and arbitrary initial and terminal conditions. Even if no general analytical solution is known for the uncontrolled dynamics, the optimally controlled system can be solved analytically. While the linearizing assumption introduced in Section 2.3 applies only to exactly realizable desired trajectories, its applicability is extended here to arbitrary desired trajectories. Thus, we proved that linear structures underlying nonlinear optimal trajectory tracking are possible.

The analytical treatment is based on a reinterpretation of a singular optimal control problem as a singularly perturbed system of differential equations. This reinterpretation is valid for all optimal control problems with affine control signals. In the light of this reinterpretation, it is now possible to understand the role of ϵ more clearly. In particular, the behavior of unregularized optimal trajectory tracking in many affine control systems can be outlined, even if they do not satisfy the linearizing assumption from Section 4.4.9.

For all $\epsilon > 0$, the dynamics of an optimal control system takes place in the extended state space, i.e., in the combined space of state $\boldsymbol{x}$ and co-state $\boldsymbol{\lambda}$ of dimension $2n$, with n being the number of state space components. For $\epsilon = 0$, the dynamics is restricted by $2p$ algebraic equations to a hypersurface of dimension $2(n-p)$, with p being the number of independent control signals. For all times except at the beginning and the end of the time interval, the system is evolving on the so-called *singular surface* (Bryson and Ho, 1975). At the initial time, a kick in form of a Dirac delta function mediated by the control signal induces an instantaneous transition from the initial state onto the singular surface. Similarly, at the terminal time, a delta-like kick induces an instantaneous transition from the singular surface to the terminal state. These instantaneous transitions render certain state components discontinuous at the initial and terminal time, respectively. For $\epsilon > 0$, the discontinuities of the state are smoothed out in form of boundary layers, i.e., continuous transition regions with a slope controlled by the value of ϵ. The control signals are finite and exhibit a sharp peak at the time domain boundaries with an amplitude inversely proportional to ϵ.

The general picture of the behavior of unregularized optimal control problems clearly explains the necessity of a regularization term in the cost functional $\mathcal{J}$. While the behavior for $\epsilon = 0$ is relatively easy to understand and determined by simpler equations than for $\epsilon > 0$, the result is mathematically inconvenient. For $\epsilon = 0$, it is not possible to find a solution for the optimal controlled state trajectory in terms of con-

tinuous functions. Even worse, the solution for the control signal must be expressed in terms of the Dirac delta function, i.e., in terms of distributions. Throughout this thesis, no attention is paid to the function spaces to which the controlled state trajectory and the control signal belongs. Everything is assumed to be sufficiently well behaved. However, the analytical treatment for $\epsilon = 0$ leads right to the importance of such questions. A mathematically more precise characterization of the different function spaces involved in the problems of optimal trajectory tracking for $\epsilon = 0$ and $\epsilon > 0$ is desirable.

Here, we derived perturbative solutions for small ϵ. Note, however, that finding such solutions is only a first step in a mathematically rigorous perturbative treatment. In a second step, existence and uniqueness of the outer and relevant inner equations together with their initial, terminal, and matching conditions must be established. Third, the reliability of the approximate result must be demonstrated. This is usually done by estimates in form of rigorous inequalities which determine how much the approximate solution deviates from the exact result for a given value of ϵ. Another point deserving more mathematical rigor concerns the linearizing assumption. Here, we showed that the linearizing assumption is a sufficient condition for a linear structure of optimal trajectory tracking. The question arises if it is also necessary. Other classes of affine control systems which violate the linearizing assumption but exhibit an underlying linear structure may exist.

Due to the limited resolution in numerical simulations, it is at least difficult, if not impossible, to find a faithful numerical representation of the solution to optimal trajectory tracking for $\epsilon = 0$. To ensure a solution to an optimal control problem in terms of numerically treatable functions, a finite value of ϵ is indispensable. For the two-dimensional dynamical systems of Section 4.1, the width of the boundary layers is directly proportional to the value of ϵ. Thus, a temporal resolution Δt smaller than ϵ, $\Delta t < \epsilon$, will not be able to numerically resolve these boundary layers, and is likely to lead to large numerical errors. Indeed, comparing a numerical result obtained for $\Delta t = \epsilon$ with its analytical counterpart reveals that the largest differences occur in the boundary layer regions, see Fig. 4.3 of Example 4.2 in Section 4.2. Note that the initial boundary layer plays an important role for the future time evolution of the system. An erroneous computation of this transition region leads to a perturbed initial value on the singular surface. This is a problem if the system is sensitive with respect to perturbations of the initial conditions. However, here the problem can be more severe due to the time dependence of the singular surface. A desired trajectory $\boldsymbol{x}_d(t)$ changing rapidly during the initial transient is likely to cause a rapidly changing singular surface, and an erroneous computation of the initial boundary layer might lead to a different singular surface altogether.

Combining analytical approximations with numerical methods can be fruitful for many applications. Analytical solutions for optimal control, even if only approximately valid, can provide a suitable initial guess for iterative optimal control algorithms, and result in a considerable decrease of computational cost. Imaginable are applications to real time computations of optimal control, which is still an ambitious

task even with modern-day fast computers. Furthermore, analytical approximations can be used to test the accuracy of numerical optimal control algorithms and estimate errors caused by discretization.

In technical application and experiments, it is impossible to generate diverging control signals, and in general, an experimental realization of unregularized optimal control systems is impossible. Nevertheless, understanding the behavior of the control system in the limit $\epsilon \to 0$ can be very useful for applications. For example, to avoid any steep transitions and large control amplitudes, one can exploit the knowledge about the initial conditions for the singular surface. If the initial state of the system can be prepared, the initial state could be chosen to lie on the singular surface. Thereby, any initial steep transitions can be prevented, or at least minimized. Furthermore, if the initial state cannot be prepared, it might still be possible to design the desired trajectory $\boldsymbol{x}_d(t)$ such that the initial state lies on the singular surface. This is only one example how an analytical solution can be utilized for the planning of desired trajectories. Another example is the discovery from Example 4.3 that the desired velocity over time can only be controlled up to a constant shift for mechanical control systems in one spatial dimension. In general, analytical solutions of optimal trajectory tracking for arbitrary desired trajectory $\boldsymbol{x}_d(t)$ enable to compare the performance of controlled state trajectories for different choices of desired trajectories. This is useful if the desired trajectory is not entirely fixed by the problem setting but exhibits some degrees of freedom. In a second step, these can be optimized with respect to other aspects as e.g. the control amplitude. Such a procedure is nearly impossible for numerical optimal control due to the computational cost of numerical algorithms.

It is clear that the class of optimal control systems with underlying linear structure is much smaller than the class of feedback linearizable systems. The reason is that the coupled state and co-state equations are more complex, and there are many more sources for nonlinearity. Consider the necessary optimality conditions for optimal trajectory tracking,

$$\mathbf{0} = \epsilon^2 \boldsymbol{u}(t) + \boldsymbol{\mathcal{B}}^T \boldsymbol{\lambda}(t), \tag{4.393}$$

$$\dot{\boldsymbol{x}}(t) = \boldsymbol{R}(\boldsymbol{x}(t)) + \boldsymbol{\mathcal{B}}\boldsymbol{u}(t), \tag{4.394}$$

$$-\dot{\boldsymbol{\lambda}}(t) = \nabla \boldsymbol{R}^T(\boldsymbol{x}(t))\,\boldsymbol{\lambda}(t) + \boldsymbol{\mathcal{S}}(\boldsymbol{x}(t) - \boldsymbol{x}_d(t)). \tag{4.395}$$

The controlled state equation is coupled with the adjoint equation via the transposed Jacobian of $\boldsymbol{R}$. Additionally, the inhomogeneity $\boldsymbol{\mathcal{S}}(\boldsymbol{x}(t) - \boldsymbol{x}_d(t))$ of the adjoint equation depends linearly on $\boldsymbol{x}$, and any nonlinear transformation of the state results in an inhomogeneity depending nonlinearly on $\boldsymbol{x}$. Applying a nonlinear state transformation, as it is often required by feedback linearization, leads to new nonlinearities in the adjoint equation. Thus, optimal control systems cannot fully benefit from the linear structure underlying feedback linearizable systems. The class of exactly linear optimal control systems is certainly much smaller than the class of feedback linearizable systems.

An important problem is the impact of noise on optimal control. Fundamental results exist for linear optimal control. The standard problem of linear optimal feedback control is the so-called linear-quadratic regulator, a linear controlled state equation together with a cost function quadratic in the state. The linear-quadratic-Gaussian control problem considers the linear-quadratic regulator together with additive Gaussian white noise in the state equation as well as for state measurements (Bryson and Ho, 1975). The discovery of linear structures underlying nonlinear trajectory tracking might enable a similar investigation for control systems satisfying the linearizing assumption. In this context, we mention Ref. (Kappen, 2005) which presents a linear theory for the control of nonlinear stochastic systems. The approach in (Kappen, 2005) relies on an exact linearization of the Hamilton-Jacobi-Bellman equation for stochastic systems by a Cole-Hopf transform. However, the method in (Kappen, 2005) is restricted to systems with identical numbers of control signals and state components, $n = p$.

4.5.2. Weak and strong coupling

Here, optimal trajectory tracking is characterized for the whole range of the regularization parameter $\epsilon \geq 0$. Consider a system which satisfies the linearizing assumption,

$$\boldsymbol{\mathcal{Q}}\boldsymbol{R}(\boldsymbol{x}) = \boldsymbol{\mathcal{Q}}\boldsymbol{\mathcal{A}}\boldsymbol{x} + \boldsymbol{\mathcal{Q}}\boldsymbol{b}, \tag{4.396}$$

such that the control signals $\boldsymbol{u}(t)$ act on nonlinear state equations. Due to the special structure defined by Eq. (4.396), the control is able to counteract the nonlinearity. In general, it is able to do so only if it is allowed to have an arbitrarily large amplitude. The amplitude of the control signal is closely related to the value of the regularization parameter ϵ. In the cost functional for optimal trajectory tracking,

$$\mathcal{J} = \frac{\alpha}{2}\int_{t_0}^{t_1} dt\,(\boldsymbol{x}(t) - \boldsymbol{x}_d(t))^2 + \frac{\epsilon^2}{2}\int_{t_0}^{t_1} dt\,(\boldsymbol{u}(t))^2, \tag{4.397}$$

the regularization term $\sim \epsilon^2$ penalizes large control signals. Depending on the value of ϵ, different regimes can be identified. Clearly, in the limit of $\epsilon \to \infty$, any non-vanishing control signal leads to a diverging value of $\mathcal{J}$. Thus, the limit $\epsilon \to \infty$ implies a vanishing control signal, $\boldsymbol{u}(t) \equiv \boldsymbol{0}$, and corresponds to the uncontrolled system. If ϵ is much larger than 1 but finite, $\epsilon \gg 1$, the control is allowed to have a small maximum amplitude. This regime can be regarded as the weak coupling limit. The nonlinearity $\boldsymbol{R}(\boldsymbol{x})$ dominates the system dynamics even if the system satisfies the linearizing assumption Eq. (4.396). For decreasing values of ϵ, the control exerts a growing influence on the system, until the regime with $1 \gg \epsilon > 0$ is reached where the control dominates over the nonlinearity. If ϵ vanishes identically, $\epsilon = 0$, the control is given by

$$\boldsymbol{u}(t) = \boldsymbol{\mathcal{B}}^{+}(\boldsymbol{x}(t))\,(\dot{\boldsymbol{x}}(t) - \boldsymbol{R}(\boldsymbol{x}(t))). \tag{4.398}$$

In general, without any constraints on the state $\boldsymbol{x}$, the nonlinearity $\boldsymbol{R}(\boldsymbol{x})$ can attain arbitrarily large values. Only if the control is allowed to attain arbitrarily large values as well, it is able to counteract an arbitrary nonlinearity $\boldsymbol{R}$ evaluated at an arbitrary state value $\boldsymbol{x}$. Only for $\epsilon = 0$, one can expect an exactly linear behavior of nonlinear optimal control systems independent of the nonlinearity $\boldsymbol{R}$.

Indeed, the analytical results indicate that the control signal scales as $1/\epsilon$. Although the analytical results are only valid for small ϵ, this corroborates the above considerations for the entire range of values of ϵ. In view of the underlying linear structure of a nonlinear optimal control for $\epsilon = 0$, one might ask if something can be learned about the nonlinear uncontrolled system by analyzing the linear controlled problem? The answer is clearly no, because the limit of an uncontrolled system is the opposite limit of an arbitrarily strongly controlled system assumed for the perturbative treatment.

This foregoing reasoning might explain why methods like feedback linearization, which exploit an underlying linear structure of controlled systems, are relatively unfamiliar in the nonlinear dynamics community and among physicists in general. Physicists tend to approach controlled systems from the viewpoint of uncontrolled systems. Having understood the manifold of solutions to the uncontrolled system, which is the traditional topic of nonlinear dynamics, the natural approach to controlled systems is to regard the control as a perturbation. This corresponds to the weak coupling limit mentioned above. Treating controlled systems in this limit is often sufficient to discuss stabilization of unstable attractors and similar topics. Such control tasks can often be achieved with non-invasive control signals. Usually, small control amplitudes are technically more feasible, and generally preferred over large control amplitudes. On the downside, concentrating solely on the weak coupling limit misses the fact that many nonlinear control systems, as e.g. feedback-linearizable systems, have an underlying linear structure. Because basically all exact linearizations of control systems work by transforming the control such that it cancels the nonlinearity, this underlying linear structure can only be exploited in the strong coupling limit and in the absence of constraints for the control. Any a priori assumptions about the maximum value of the control amplitude, enforced by a regularization term or inequality constraints in case of optimal control, destroy the underlying linear structure.

However, the strong coupling limit does not always imply very large or even diverging control amplitudes. After having obtained the solutions for the control signal as well as the controlled state trajectory, it is possible to give a posteriori estimates on the maximum control amplitude. Depending on the desired trajectories, initial conditions, and system dynamics, these a posteriori estimates can be comparatively small. The weak coupling limit only refers to a priori assumptions about the maximum control amplitude, and excludes or penalizes large control amplitudes from the very beginning.

5. Control of reaction-diffusion systems

Reaction-diffusion systems model phenomena from a large variety of fields. Examples are chemical systems (Kapral and Showalter, 1995; Epstein and Pojman, 1998), action potential propagation in the heart and neurons (Keener and Sneyd, 2008a,b), population dynamics (Murray, 2007, 2011), vegetation patterns (von Hardenberg et al., 2001), and the motility of crawling cells (Ziebert et al., 2011; Ziebert and Aranson, 2013; Löber et al., 2014; Aranson et al., 2014; Löber et al., 2015), to name only a few. These systems possess a rich phenomenology of solutions, ranging from homogeneous stable steady states, phase waves, Turing patterns, stationary localized and labyrinthine patterns, traveling, rotating and scroll waves to fully developed spatio-temporal turbulence (Turing, 1952; Cross and Hohenberg, 1993; Hagberg and Meron, 1994; Kuramoto, 2003; Vanag and Epstein, 2007). Due to the complexity and the nonlinearity of the underlying evolution equations, their theoretical investigation relies heavily on numerical simulations. However, more complex patterns can often be understood as being assembled of simple "building blocks" as traveling fronts and pulses. A solitary pulse in the FHN model in one spatial dimension can be considered as being built of two propagating interfaces separating the excited from the refractory state. These interfaces are front solutions to a simpler reaction-diffusion system. Similarly, many two-dimensional shapes as e.g. spiral waves can be approximated as consisting of appropriately shifted one-dimensional pulse profiles (Tyson and Keener, 1988; Zykov, 1988; Pismen, 2006; Löber and Engel, 2013; Mikhailov, 2011). In many cases, the simplified equations allow the inclusion of additional effects as e.g. spatial heterogeneities (Löber, 2009; Alonso et al., 2010; Löber et al., 2012), noise (Schimansky-Geier et al., 1983; Engel, 1985), or curved boundaries (Engel and Ebeling, 1987; Martens et al., 2015).

The control of patterns in reaction-diffusion system has received the attention of many researchers in the past (Mikhailov and Showalter, 2006; Vanag and Epstein, 2008). Due to their complexity, it makes sense to develop first a detailed understanding of the control of simple solutions as e.g. solitary excitation pulses. A particularly simple but still general control task is position control of traveling waves. The position of a traveling wave is shifted according to a prescribed trajectory in position space, called the protocol of motion. Simultaneously, the wave profile is kept as close as possible to the uncontrolled wave profile. An example of open loop control in this spirit is the dragging of chemical pulses of adsorbed CO during heterogeneous catalysis on platinum single crystal surfaces (Wolff et al., 2003b). In

experiments with an addressable catalyst surface, the pulse velocity was controlled by a laser beam creating a movable localized temperature heterogeneity, resulting in a V-shaped wave pattern (Wolff et al., 2001, 2003a). Theoretical studies of dragging one-dimensional chemical fronts or phase interfaces by anchoring it to a movable parameter heterogeneity can be found in (Nistazakis et al., 2002; Malomed et al., 2002; Kevrekidis et al., 2004). While these approaches assume a fixed spatial profile of the control signal and vary only its location, the method developed in (Löber and Engel, 2014) determines the profile, amplitude, and location of the control signal by solving an inverse problem for the position over time of controlled traveling waves. This control solution is close to the solution of an appropriately formulated optimal control problem. Furthermore, an extension allows the investigation of the stability of controlled traveling waves (Löber, 2014), which can never be taken for granted in open loop control systems. A modification of the method provides shaping of wave patterns in two-dimensional reaction-diffusion systems by shifting the position of a trajectory outlining the pattern (Löber et al., 2014). See also (Löber et al., 2014) for a discussion of experimental realizations. An approach similarly aiming at the position of wave patterns is the forcing of spiral waves with temporally periodic and spatially homogeneous control signals. This can be utilized to guide a meandering spiral wave tip along a wide range of open and closed hypocycloidal trajectories (Steinbock et al., 1993; Zykov et al., 1994).

Position control can be tackled by feedback control as well. In experiments with spiral waves in the photosensitive Belousov-Zhabotinsky reaction (Krug et al., 1990), the spiral wave core is steered around obstacles using feedback signals obtained from wave activity measured at detector points, along detector lines, or in a spatially extended control domain (Zykov et al., 2004; Zykov and Engel, 2004; Schlesner et al., 2008). Two feedback loops were used to guide wave segments along pre-given trajectories (Sakurai et al., 2002). Furthermore, feedback-mediated control loops are employed in order to stabilize unstable patterns such as plane waves undergoing transversal instabilities (Molnos et al., 2015), unstable traveling wave segments (Mihaliuk et al., 2002), or rigidly rotating unstable spiral waves in the regime of stable meandering spiral waves (Schlesner et al., 2006). Another strategy is control by imposed geometric constraints such as no-flux boundaries (Paulau et al., 2013) or heterogeneities (Luther et al., 2011).

While feedback control and external forcing of reaction-diffusion system has received much attention, optimal control of these systems remains largely unexplored, at least within the physics community. One reason lies in the computational cost involved in numerical approaches to optimal open loop control of PDEs which restricts numerical investigations to relatively small spatial domains and short time intervals. Even worse, optimal feedback control of PDEs becomes almost intractable. The reason is the curse of dimensionality (Bellman, 2003). For an n-dimensional dynamical system, the Hamilton-Jacobi-Bellman equation for optimal feedback control is a PDE on an n-dimensional domain, see the discussion at the beginning of Section 4.3. For a controlled PDE, which can be regarded as a dynamical system with

$n \to \infty$ dimensions, the corresponding Hamilton-Jacobi-Bellman equation is a PDE on a domain with $n \to \infty$ dimensions.

In view of the numerical difficulties, the analytical approach pursued in this thesis has some benefits when compared with purely numerical methods, and can be used to obtain solutions to optimal control for a number of systems with relative ease. In Section 5.1, the formalism based on projectors is modified and applied to spatio-temporal systems. The controlled state equation is split up in two equations in Section 5.2, and exactly realizable distributions are introduced as the spatio-temporal analogue of exactly realizable trajectories in Section 5.3. As an important application, the position control of traveling waves is discussed in Section 5.4. This chapter concludes with a discussion and outlook in Section 5.5.

5.1. Formalism

In this section, the formalism developed in Chapter 2 is modified and applied to spatio-temporal systems. The emphasis lies on distributed controls, i.e., the control signal is allowed to depend on space and time. Often, the control cannot act everywhere in position space. For example, it might be possible to let a control act at or close to the boundaries, but it is impossible to reach the interior of the domain. These restrictions can be accounted for by a modification of the projectors $\boldsymbol{\mathcal{P}}$ and $\boldsymbol{\mathcal{Q}}$.

Let the position vector $\boldsymbol{r}$ in N spatial dimensions be

$$\boldsymbol{r} = \begin{pmatrix} r_1, & r_2, \ldots, r_N \end{pmatrix}^T . \tag{5.1}$$

The spatial domain is denoted by $\Omega \subset \mathbb{R}^N$, and its boundary is $\Gamma = \partial\Omega \subset \mathbb{R}^N$. Let $\boldsymbol{\chi}(\boldsymbol{r})$ be a diagonal $n \times n$ matrix of characteristic functions,

$$\boldsymbol{\chi}(\boldsymbol{r}) = \begin{pmatrix} \chi_1(\boldsymbol{r}) & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \chi_n(\boldsymbol{r}) \end{pmatrix} . \tag{5.2}$$

The characteristic function $\chi_i(\boldsymbol{r})$ only attains the values zero or one,

$$\chi_i(\boldsymbol{r}) = \begin{cases} 1, & \boldsymbol{r} \in A_i, \\ 0, & \boldsymbol{r} \in \Omega \setminus A_i, \end{cases} \quad i \in \{1, \ldots, n\} . \tag{5.3}$$

The characteristic functions $\chi_i(\boldsymbol{r})$ do account for the case that a control signal might only act in a restricted region of space and not in the full domain Ω. The region in which a control signal acts on the i-th state component is denoted by A_i. The total spatial region affected by control is $A = \bigcup_{i=1}^{N} A_i$, and no control acts in region $\Omega \setminus A$ outside of A. If all state components are controlled in the same region

$A_1 = A_2 = \cdots = A_n$ of space, then $\boldsymbol{\chi}(\boldsymbol{r})$ simplifies to a multiple of the identity matrix $\boldsymbol{\chi}(\boldsymbol{r}) = \chi(\boldsymbol{r})\,\mathbf{1}$, with a scalar characteristic function χ. If additionally, the control acts everywhere in the spatial domain Ω, then $\boldsymbol{\chi}(\boldsymbol{r}) = \mathbf{1}$. The matrix $\boldsymbol{\chi}(\boldsymbol{r})$ is a space dependent projector on the state space. It is idempotent,

$$\boldsymbol{\chi}(\boldsymbol{r})\,\boldsymbol{\chi}(\boldsymbol{r}) = \boldsymbol{\chi}(\boldsymbol{r}), \tag{5.4}$$

and symmetric

$$\boldsymbol{\chi}^T(\boldsymbol{r}) = \boldsymbol{\chi}(\boldsymbol{r}). \tag{5.5}$$

The projector $\boldsymbol{\psi}$ complementary to $\boldsymbol{\chi}$ is defined as

$$\boldsymbol{\psi}(\boldsymbol{r}) = \mathbf{1} - \boldsymbol{\chi}(\boldsymbol{r}), \tag{5.6}$$

such that $\boldsymbol{\psi}(\boldsymbol{r})\,\boldsymbol{\chi}(\boldsymbol{r}) = \mathbf{0}$.

An affine controlled reaction-diffusion system for the n-component state vector

$$\boldsymbol{x}(\boldsymbol{r},t) = \left(\; x_1(\boldsymbol{r},t),\quad \ldots,\quad x_n(\boldsymbol{r},t) \;\right) \tag{5.7}$$

with the p-component vector of distributed control signals

$$\boldsymbol{u}(\boldsymbol{r},t) = \left(\; u_1(\boldsymbol{r},t),\quad \ldots,\quad u_p(\boldsymbol{r},t) \;\right) \tag{5.8}$$

is

$$\partial_t \boldsymbol{x}(\boldsymbol{r},t) = \boldsymbol{\mathcal{D}} \triangle \boldsymbol{x}(\boldsymbol{r},t) + \boldsymbol{R}(\boldsymbol{x}(\boldsymbol{r},t)) + \boldsymbol{\chi}(\boldsymbol{r})\,\boldsymbol{\mathcal{B}}(\boldsymbol{x}(\boldsymbol{r},t))\,\boldsymbol{u}(\boldsymbol{r},t). \tag{5.9}$$

Here, $\boldsymbol{\mathcal{D}}$ is an $n \times n$ diagonal matrix of constant diffusion coefficients and $\triangle$ denotes the Laplacian which, in Cartesian coordinates, assumes the form $\triangle = \sum_{i=1}^{N} \dfrac{\partial^2}{\partial r_i^2}$. For simplicity, an isotropic medium is considered. The $n \times p$ coupling matrix $\boldsymbol{\mathcal{B}}(\boldsymbol{x})$ is assumed to have full column rank, $\operatorname{rank}(\boldsymbol{\mathcal{B}}(\boldsymbol{x})) = p$, for all $\boldsymbol{x}$. To shorten the notation, the $n \times p$ matrix

$$\boldsymbol{\mathcal{B}}(\boldsymbol{x},\boldsymbol{r}) = \boldsymbol{\chi}(\boldsymbol{r})\,\boldsymbol{\mathcal{B}}(\boldsymbol{x}) \tag{5.10}$$

is introduced. For reaction-diffusion systems in finite domains, Eq. (5.9) is supplemented with appropriate boundary conditions. A common choice are homogeneous Neumann or no flux boundary conditions

$$\boldsymbol{n}^T(\boldsymbol{r})\,(\boldsymbol{\mathcal{D}} \nabla \boldsymbol{x}(\boldsymbol{r},t)) = \mathbf{0},\ \boldsymbol{r} \in \Gamma. \tag{5.11}$$

Here, the N-component vector $\boldsymbol{n}(\boldsymbol{r})$ is the vector normal to the boundary Γ. If the diffusion coefficient for a certain component vanishes, the boundary condition for this component is trivially satisfied.

In principle, additional control signals acting on the domain boundary Γ can be introduced. In case of Neumann boundary conditions, this corresponds to an inhomogeneity on the right hand side of Eq. (5.11) prescribing the flux of state components across Γ (Theißen, 2006). See (Lebiedz and Brandt-Pollmann, 2003) how a desired stationary concentration profile is enforced in a reaction-diffusion system by boundary control. Although such control schemes are important for applications, the discussion here is restricted to distributed controls, i.e., spatio-temporal control signals acting inside the spatial domain. Other possible boundary conditions for reaction-diffusion systems are Dirichlet or periodic boundary conditions. Finally, the initial condition for Eq. (5.9) is given by

$$\boldsymbol{x}\left(\boldsymbol{r}, t_0\right) = \boldsymbol{x}_0\left(\boldsymbol{r}\right). \tag{5.12}$$

5.2. Split up the state equation

Similar as in earlier chapters, the control signal can be expressed in terms of the controlled state $\boldsymbol{x}\left(\boldsymbol{r}, t\right)$ as

$$\boldsymbol{u}\left(\boldsymbol{r}, t\right) = \boldsymbol{\mathcal{B}}^{+}\left(\boldsymbol{x}\left(\boldsymbol{r}, t\right), \boldsymbol{r}\right)\left(\partial_t \boldsymbol{x}\left(\boldsymbol{r}, t\right) - \boldsymbol{\mathcal{D}} \triangle \boldsymbol{x}\left(\boldsymbol{r}, t\right) - \boldsymbol{R}\left(\boldsymbol{x}\left(\boldsymbol{r}, t\right)\right)\right). \tag{5.13}$$

Here, $\boldsymbol{\mathcal{B}}^{+}\left(\boldsymbol{x}, \boldsymbol{r}\right)$ denotes the $p \times n$ Moore-Penrose pseudo inverse of the matrix $\boldsymbol{\mathcal{B}}\left(\boldsymbol{x}, \boldsymbol{r}\right)$,

$$\boldsymbol{\mathcal{B}}^{+}\left(\boldsymbol{x}, \boldsymbol{r}\right) = \left(\boldsymbol{\chi}\left(\boldsymbol{r}\right)\boldsymbol{\mathcal{B}}\left(\boldsymbol{x}\right)\right)^{+}. \tag{5.14}$$

In contrast to earlier chapters, no closed form expression can be given for the pseudo inverse $\boldsymbol{\mathcal{B}}^{+}\left(\boldsymbol{x}, \boldsymbol{r}\right)$ in the general case. The reason is that the $p \times p$ matrix $\boldsymbol{\mathcal{B}}^{T}\left(\boldsymbol{x}, \boldsymbol{r}\right)\boldsymbol{\mathcal{B}}\left(\boldsymbol{x}, \boldsymbol{r}\right) = \boldsymbol{\mathcal{B}}^{T}\left(\boldsymbol{x}\right)\boldsymbol{\chi}\left(\boldsymbol{r}\right)\boldsymbol{\mathcal{B}}\left(\boldsymbol{x}\right)$ does not necessarily have rank p and can therefore not be inverted. In general, the rank of $\boldsymbol{\mathcal{B}}^{T}\left(\boldsymbol{x}\right)\boldsymbol{\chi}\left(\boldsymbol{r}\right)\boldsymbol{\mathcal{B}}\left(\boldsymbol{x}\right)$ depends on the spatial coordinate $\boldsymbol{r}$. Nevertheless, the Moore-Penrose pseudo inverse of $\boldsymbol{\mathcal{B}}^{+}\left(\boldsymbol{x}, \boldsymbol{r}\right)$ does always exist and can be computed numerically with the help of singular value decomposition, for example. In some special but important cases, explicit expressions for the pseudo inverse $\boldsymbol{\mathcal{B}}^{+}\left(\boldsymbol{x}, \boldsymbol{r}\right)$ can be given. Note that $\boldsymbol{\chi}\left(\boldsymbol{r}\right)$, being a symmetric projector, is its own pseudo inverse, i.e.,

$$\boldsymbol{\chi}^{+}\left(\boldsymbol{r}\right) = \boldsymbol{\chi}\left(\boldsymbol{r}\right). \tag{5.15}$$

If all state components are affected in the same region of space such that $\boldsymbol{\chi}\left(\boldsymbol{r}\right) = \chi\left(\boldsymbol{r}\right)\mathbf{1}$, then the pseudo inverse is given by

$$\boldsymbol{\mathcal{B}}^{+}\left(\boldsymbol{x}, \boldsymbol{r}\right) = \chi\left(\boldsymbol{r}\right)\boldsymbol{\mathcal{B}}^{+}\left(\boldsymbol{x}\right) = \chi\left(\boldsymbol{r}\right)\left(\boldsymbol{\mathcal{B}}^{T}\left(\boldsymbol{x}\right)\boldsymbol{\mathcal{B}}\left(\boldsymbol{x}\right)\right)^{-1}\boldsymbol{\mathcal{B}}^{T}\left(\boldsymbol{x}\right). \tag{5.16}$$

If the number of independent control signals equals the number of state components, $n = p$, such that $\boldsymbol{\mathcal{B}}\left(\boldsymbol{x}\right)$ is invertible, then the pseudo inverse is

$$\boldsymbol{\mathcal{B}}^{+}\left(\boldsymbol{x}, \boldsymbol{r}\right) = \boldsymbol{\mathcal{B}}^{-1}\left(\boldsymbol{x}\right)\boldsymbol{\chi}^{+}\left(\boldsymbol{r}\right) = \boldsymbol{\mathcal{B}}^{-1}\left(\boldsymbol{x}\right)\boldsymbol{\chi}\left(\boldsymbol{r}\right). \tag{5.17}$$

Finally, if the $p \times p$ matrix $\boldsymbol{\mathcal{B}}^T(\boldsymbol{x})\,\boldsymbol{\chi}(\boldsymbol{r})\,\boldsymbol{\mathcal{B}}(\boldsymbol{x})$ has full rank p for all values of $\boldsymbol{r} \in \Omega$ and for all states $\boldsymbol{x}$, the pseudo inverse is given by

$$\boldsymbol{\mathcal{B}}^+(\boldsymbol{x},\boldsymbol{r}) = \left(\boldsymbol{\mathcal{B}}^T(\boldsymbol{x})\,\boldsymbol{\chi}(\boldsymbol{r})\,\boldsymbol{\mathcal{B}}(\boldsymbol{x})\right)^{-1}\boldsymbol{\mathcal{B}}^T(\boldsymbol{x})\,\boldsymbol{\chi}(\boldsymbol{r}). \tag{5.18}$$

Note that for any matrix $\boldsymbol{\mathcal{A}}$, its Moore-Penrose pseudo inverse can also be expressed as (Albert, 1972)

$$\boldsymbol{\mathcal{A}}^+ = \left(\boldsymbol{\mathcal{A}}^T\boldsymbol{\mathcal{A}}\right)^+\boldsymbol{\mathcal{A}}^T, \tag{5.19}$$

such that $\boldsymbol{\mathcal{B}}^+(\boldsymbol{x},\boldsymbol{r})$ can be written in the form

$$\begin{aligned}\boldsymbol{\mathcal{B}}^+(\boldsymbol{x},\boldsymbol{r}) &= \left(\boldsymbol{\mathcal{B}}^T(\boldsymbol{x},\boldsymbol{r})\,\boldsymbol{\mathcal{B}}(\boldsymbol{x},\boldsymbol{r})\right)^+\boldsymbol{\mathcal{B}}^T(\boldsymbol{x},\boldsymbol{r}) \\ &= \left(\boldsymbol{\mathcal{B}}^T(\boldsymbol{x})\,\boldsymbol{\chi}(\boldsymbol{r})\,\boldsymbol{\mathcal{B}}(\boldsymbol{x})\right)^+\boldsymbol{\mathcal{B}}^T(\boldsymbol{x})\,\boldsymbol{\chi}(\boldsymbol{r}).\end{aligned} \tag{5.20}$$

Together with the projector property of $\boldsymbol{\chi}(\boldsymbol{r})$, Eq. (5.20) yields the identity

$$\boldsymbol{\mathcal{B}}^+(\boldsymbol{x},\boldsymbol{r})\,\boldsymbol{\chi}(\boldsymbol{r}) = \boldsymbol{\mathcal{B}}^+(\boldsymbol{x},\boldsymbol{r}), \qquad \boldsymbol{\mathcal{B}}^+(\boldsymbol{x},\boldsymbol{r})\,\boldsymbol{\psi}(\boldsymbol{r}) = \mathbf{0}. \tag{5.21}$$

Consequently, the control can be written as

$$\boldsymbol{u}(\boldsymbol{r},t) = \boldsymbol{\mathcal{B}}^+(\boldsymbol{x}(\boldsymbol{r},t),\boldsymbol{r})\,\boldsymbol{\chi}(\boldsymbol{r})\left(\partial_t\boldsymbol{x}(\boldsymbol{r},t) - \boldsymbol{\mathcal{D}}\triangle\boldsymbol{x}(\boldsymbol{r},t) - \boldsymbol{R}(\boldsymbol{x}(\boldsymbol{r},t))\right). \tag{5.22}$$

As could be expected intuitively, evaluating this expression at a position $\boldsymbol{r} \in \Omega \setminus A$ outside the region A affected by control yields a vanishing control signal,

$$\boldsymbol{u}(\boldsymbol{r},t) = \mathbf{0},\ \boldsymbol{r} \in \Omega \setminus A. \tag{5.23}$$

Using expression (5.13) for the control signal in the controlled reaction-diffusion system, Eq. (5.9), yields

$$\begin{aligned}\partial_t\boldsymbol{x}(\boldsymbol{r},t) = {}& \boldsymbol{\mathcal{Q}}(\boldsymbol{x}(\boldsymbol{r},t),\boldsymbol{r})\left(\boldsymbol{\mathcal{D}}\triangle\boldsymbol{x}(\boldsymbol{r},t) + \boldsymbol{R}(\boldsymbol{x}(\boldsymbol{r},t))\right) \\ &+ \boldsymbol{\mathcal{P}}(\boldsymbol{x}(\boldsymbol{r},t),\boldsymbol{r})\,\partial_t\boldsymbol{x}(\boldsymbol{r},t),\end{aligned} \tag{5.24}$$

or

$$\mathbf{0} = \boldsymbol{\mathcal{Q}}(\boldsymbol{x}(\boldsymbol{r},t),\boldsymbol{r})\left(\partial_t\boldsymbol{x}(\boldsymbol{r},t) - \boldsymbol{\mathcal{D}}\triangle\boldsymbol{x}(\boldsymbol{r},t) - \boldsymbol{R}(\boldsymbol{x}(\boldsymbol{r},t))\right). \tag{5.25}$$

Equation (5.25) is the constraint equation for reaction-diffusion systems. The two complementary $n \times n$ projectors $\boldsymbol{\mathcal{P}}$ and $\boldsymbol{\mathcal{Q}}$ are given by

$$\boldsymbol{\mathcal{P}}(\boldsymbol{x},\boldsymbol{r}) = \boldsymbol{\mathcal{B}}(\boldsymbol{x},\boldsymbol{r})\,\boldsymbol{\mathcal{B}}^+(\boldsymbol{x},\boldsymbol{r}) = \boldsymbol{\chi}(\boldsymbol{r})\,\boldsymbol{\mathcal{B}}(\boldsymbol{x})\,\boldsymbol{\mathcal{B}}^+(\boldsymbol{x},\boldsymbol{r})\,\boldsymbol{\chi}(\boldsymbol{r}), \tag{5.26}$$

$$\boldsymbol{\mathcal{Q}}(\boldsymbol{x},\boldsymbol{r}) = \mathbf{1} - \boldsymbol{\mathcal{P}}(\boldsymbol{x},\boldsymbol{r}). \tag{5.27}$$

For general spatio-temporal control systems, the projectors $\boldsymbol{\mathcal{P}}$ and $\boldsymbol{\mathcal{Q}}$ do not only depend the state $\boldsymbol{x}$, but also on the position $\boldsymbol{r}$. Acting with projectors $\boldsymbol{\psi}$ and $\boldsymbol{\chi}$ on $\boldsymbol{\mathcal{P}}$ and $\boldsymbol{\mathcal{Q}}$ and using also Eq. (5.20) yields the following relations,

$$\boldsymbol{\psi}(\boldsymbol{r})\,\boldsymbol{\mathcal{P}}(\boldsymbol{x},\boldsymbol{r}) = \mathbf{0}, \qquad \boldsymbol{\chi}(\boldsymbol{r})\,\boldsymbol{\mathcal{P}}(\boldsymbol{x},\boldsymbol{r}) = \boldsymbol{\mathcal{P}}(\boldsymbol{x},\boldsymbol{r}), \tag{5.28}$$

$$\boldsymbol{\psi}(\boldsymbol{r})\,\boldsymbol{\mathcal{Q}}(\boldsymbol{x},\boldsymbol{r}) = \boldsymbol{\psi}(\boldsymbol{r}), \quad \boldsymbol{\chi}(\boldsymbol{r})\,\boldsymbol{\mathcal{Q}}(\boldsymbol{x},\boldsymbol{r}) = \boldsymbol{\chi}(\boldsymbol{r}) - \boldsymbol{\mathcal{P}}(\boldsymbol{x},\boldsymbol{r}) = \boldsymbol{\mathcal{Q}}(\boldsymbol{x},\boldsymbol{r}) - \boldsymbol{\psi}(\boldsymbol{r}), \tag{5.29}$$

$$\boldsymbol{\mathcal{P}}(\boldsymbol{x},\boldsymbol{r})\,\boldsymbol{\psi}(\boldsymbol{r}) = \mathbf{0}, \qquad \boldsymbol{\mathcal{P}}(\boldsymbol{x},\boldsymbol{r})\,\boldsymbol{\chi}(\boldsymbol{r}) = \boldsymbol{\mathcal{P}}(\boldsymbol{x},\boldsymbol{r}), \tag{5.30}$$

$$\boldsymbol{\mathcal{Q}}(\boldsymbol{x},\boldsymbol{r})\,\boldsymbol{\psi}(\boldsymbol{r}) = \boldsymbol{\psi}(\boldsymbol{r}), \quad \boldsymbol{\mathcal{Q}}(\boldsymbol{x},\boldsymbol{r})\,\boldsymbol{\chi}(\boldsymbol{r}) = \boldsymbol{\chi}(\boldsymbol{r}) - \boldsymbol{\mathcal{P}}(\boldsymbol{x},\boldsymbol{r}) = \boldsymbol{\mathcal{Q}}(\boldsymbol{x},\boldsymbol{r}) - \boldsymbol{\psi}(\boldsymbol{r}). \tag{5.31}$$

The constraint equation (5.25) can be interpreted as follows. Acting with $\boldsymbol{\psi}(\boldsymbol{r})$ from the left on the constraint equation (5.25) yields

$$\mathbf{0} = \boldsymbol{\psi}(\boldsymbol{r})\left(\partial_t \boldsymbol{x}(\boldsymbol{r},t) - \boldsymbol{\mathcal{D}}\triangle\boldsymbol{x}(\boldsymbol{r},t) - \boldsymbol{R}(\boldsymbol{x}(\boldsymbol{r},t))\right). \tag{5.32}$$

Thus, outside the spatial region A affected by control, the state $\boldsymbol{x}(\boldsymbol{r},t)$ satisfies the uncontrolled reaction-diffusion equation. Acting with $\boldsymbol{\chi}(\boldsymbol{r})$ from the left on Eq. (5.25) yields an equation which is equivalent to the constraint equation for dynamical systems,

$$\mathbf{0} = \boldsymbol{\chi}(\boldsymbol{r})\,\boldsymbol{\mathcal{Q}}(\boldsymbol{x},\boldsymbol{r})\left(\partial_t \boldsymbol{x}(\boldsymbol{r},t) - \boldsymbol{\mathcal{D}}\triangle\boldsymbol{x}(\boldsymbol{r},t) - \boldsymbol{R}(\boldsymbol{x}(\boldsymbol{r},t))\right). \tag{5.33}$$

Inside the region A affected by control, p state components determine the vector of control signals $\boldsymbol{u}(t)$ while the remaining $n-p$ components are fixed by Eq. (5.33).

5.3. Exactly realizable distributions

The spatio-temporal analogues of desired trajectories in dynamical systems are called desired distributions. Exactly realizable distributions are desired distributions for which a control signal can be found such that the state $\boldsymbol{x}(\boldsymbol{r},t)$ equals the desired distribution $\boldsymbol{x}_d(\boldsymbol{r},t)$ everywhere and for all times,

$$\boldsymbol{x}(\boldsymbol{r},t) = \boldsymbol{x}_d(\boldsymbol{r},t). \tag{5.34}$$

For a desired distribution $\boldsymbol{x}_d(\boldsymbol{r},t)$ to be exactly realizable, it has to satisfy the constraint equation

$$\mathbf{0} = \boldsymbol{\mathcal{Q}}(\boldsymbol{x}_d(\boldsymbol{r},t),\boldsymbol{r})\left(\partial_t \boldsymbol{x}_d(\boldsymbol{r},t) - \boldsymbol{\mathcal{D}}\triangle\boldsymbol{x}_d(\boldsymbol{r},t) - \boldsymbol{R}(\boldsymbol{x}_d(\boldsymbol{r},t))\right). \tag{5.35}$$

Furthermore, the desired distribution $\boldsymbol{x}_d(\boldsymbol{r},t)$ must comply with the initial and boundary conditions for the state,

$$\boldsymbol{x}_d(\boldsymbol{r},t_0) = \boldsymbol{x}(\boldsymbol{r},t_0), \qquad \boldsymbol{n}^T(\boldsymbol{r})\left(\boldsymbol{\mathcal{D}}\nabla\boldsymbol{x}_d(\boldsymbol{r},t)\right) = \mathbf{0},\ \boldsymbol{r}\in\Gamma. \tag{5.36}$$

The control signal enforcing the exactly realizable distribution $\boldsymbol{x}_d(\boldsymbol{r},t)$ is given by

$$\boldsymbol{u}(\boldsymbol{r},t) = \boldsymbol{\mathcal{B}}^{+}(\boldsymbol{x}_d(\boldsymbol{r},t),\boldsymbol{r})\,\boldsymbol{\chi}(\boldsymbol{r})\,(\partial_t \boldsymbol{x}_d(\boldsymbol{r},t) - \boldsymbol{\mathcal{D}}\triangle \boldsymbol{x}_d(\boldsymbol{r},t) - \boldsymbol{R}(\boldsymbol{x}_d(\boldsymbol{r},t)))\,. \quad (5.37)$$

The proof that these assumptions lead to the desired distribution being an exact solution for the state, $\boldsymbol{x}(\boldsymbol{r},t) = \boldsymbol{x}_d(\boldsymbol{r},t)$, is analogous to the proof for dynamical systems from Section 2.2. In short, introducing $\boldsymbol{y}$ as

$$\boldsymbol{x}(\boldsymbol{r},t) = \boldsymbol{x}_d(\boldsymbol{r},t) + \boldsymbol{y}(\boldsymbol{r},t)\,, \quad (5.38)$$

and using the control signal Eq. (5.37) in the controlled state equation (5.9) yields, together with the constraint equation (5.35) and after linearization in $\boldsymbol{y}$, a linear homogeneous partial differential equation for $\boldsymbol{y}$,

$$\partial_t \boldsymbol{y}(\boldsymbol{r},t) = \boldsymbol{\mathcal{D}}\triangle \boldsymbol{y}(\boldsymbol{r},t) + (\nabla \boldsymbol{R}(\boldsymbol{x}_d(\boldsymbol{r},t)) + \boldsymbol{\mathcal{T}}(\boldsymbol{x}_d(\boldsymbol{r},t),\boldsymbol{r}))\,\boldsymbol{y}(\boldsymbol{r},t)\,. \quad (5.39)$$

The $n \times n$ matrix $\boldsymbol{\mathcal{T}}(\boldsymbol{x},\boldsymbol{r})$ is defined by

$$\boldsymbol{\mathcal{T}}(\boldsymbol{x},\boldsymbol{r})\,\boldsymbol{y} = \boldsymbol{\chi}(\boldsymbol{r})\,(\nabla \boldsymbol{\mathcal{B}}(\boldsymbol{x})\,\boldsymbol{y})\,\boldsymbol{\mathcal{B}}^{+}(\boldsymbol{x},\boldsymbol{r})\,\boldsymbol{\chi}(\boldsymbol{r})\,(\partial_t \boldsymbol{x} - \boldsymbol{\mathcal{D}}\triangle \boldsymbol{x} - \boldsymbol{R}(\boldsymbol{x}))\,. \quad (5.40)$$

Equation (5.39) is to be solved with the initial condition

$$\boldsymbol{y}(\boldsymbol{r},t_0) = \boldsymbol{x}_0(\boldsymbol{r}) - \boldsymbol{x}_d(\boldsymbol{r},t)\,. \quad (5.41)$$

If the initial state $\boldsymbol{x}_0(\boldsymbol{r})$ complies with the initial desired distribution, then $\boldsymbol{y}(\boldsymbol{r},t_0) = \boldsymbol{0}$ initially. Furthermore, if $\boldsymbol{x}(\boldsymbol{r},t)$ as well as $\boldsymbol{x}_d(\boldsymbol{r},t)$ satisfy homogeneous Neumann boundary conditions, then

$$\boldsymbol{n}^T(\boldsymbol{r})\,(\boldsymbol{\mathcal{D}}\nabla \boldsymbol{y}(\boldsymbol{r},t)) = \boldsymbol{0},\ \boldsymbol{r} \in \Gamma. \quad (5.42)$$

Consequently, $\boldsymbol{y}$ vanishes everywhere and for all times,

$$\boldsymbol{y}(\boldsymbol{r},t) \equiv \boldsymbol{0}. \quad (5.43)$$

Equation (5.39) determines the stability of exactly realizable trajectories against perturbations $\boldsymbol{y}(\boldsymbol{r},t_0) = \boldsymbol{y}_0$ of the initial conditions.

Similar as for dynamical systems, a linearizing assumption can be introduced. First of all, the projectors $\boldsymbol{\mathcal{P}}$ and $\boldsymbol{\mathcal{Q}}$ must be independent of the state $\boldsymbol{x}$,

$$\boldsymbol{\mathcal{Q}}(\boldsymbol{x},\boldsymbol{r}) = \boldsymbol{\mathcal{Q}}(\boldsymbol{r}) = \text{const.} \quad (5.44)$$

Second, the nonlinearity $\boldsymbol{R}(\boldsymbol{x})$ must satisfy the condition

$$\boldsymbol{\mathcal{Q}}(\boldsymbol{r})\,\boldsymbol{R}(\boldsymbol{x}) = \boldsymbol{\mathcal{Q}}(\boldsymbol{r})\,\boldsymbol{\mathcal{A}}\boldsymbol{x} + \boldsymbol{\mathcal{Q}}(\boldsymbol{r})\,\boldsymbol{b}, \quad (5.45)$$

such that the constraint equation (5.35) becomes linear. In principle, the projector $\boldsymbol{\mathcal{Q}}(\boldsymbol{r})$ may still depend on the position $\boldsymbol{r}$ to yield a linear constraint equation. This

might be useful if the nonlinearity $\boldsymbol{R}(\boldsymbol{x})$ exhibits an explicit dependence on space $\boldsymbol{r}$.

Equation (5.35) becomes a linear PDE for $\boldsymbol{\mathcal{Q}}(\boldsymbol{r})\,\boldsymbol{x}_d(\boldsymbol{r},t)$ with $\boldsymbol{\mathcal{P}}(\boldsymbol{r})\,\boldsymbol{x}_d(\boldsymbol{r},t)$ serving as an inhomogeneity,

$$\boldsymbol{\mathcal{Q}}(\boldsymbol{r})\,\partial_t \boldsymbol{x}_d(\boldsymbol{r},t) = \boldsymbol{\mathcal{Q}}(\boldsymbol{r})\,\boldsymbol{\mathcal{D}}\triangle \boldsymbol{x}_d(\boldsymbol{r},t) + \boldsymbol{\mathcal{Q}}(\boldsymbol{r})\,\boldsymbol{\mathcal{A}}\boldsymbol{x}_d(\boldsymbol{r},t) + \boldsymbol{\mathcal{Q}}(\boldsymbol{r})\,\boldsymbol{b}, \tag{5.46}$$

or, inserting $\mathbf{1} = \boldsymbol{\mathcal{P}}(\boldsymbol{r}) + \boldsymbol{\mathcal{Q}}(\boldsymbol{r})$ between $\boldsymbol{\mathcal{A}}$ and $\boldsymbol{x}_d$,

$$\begin{aligned}&\partial_t\left(\boldsymbol{\mathcal{Q}}(\boldsymbol{r})\,\boldsymbol{x}_d(\boldsymbol{r},t)\right) - \boldsymbol{\mathcal{Q}}(\boldsymbol{r})\,\boldsymbol{\mathcal{D}}\triangle\left(\boldsymbol{\mathcal{Q}}(\boldsymbol{r})\,\boldsymbol{x}_d(\boldsymbol{r},t)\right) - \boldsymbol{\mathcal{Q}}(\boldsymbol{r})\,\boldsymbol{\mathcal{A}}\boldsymbol{\mathcal{Q}}(\boldsymbol{r})\,\boldsymbol{x}_d(\boldsymbol{r},t)\\ =&\boldsymbol{\mathcal{Q}}(\boldsymbol{r})\,\boldsymbol{\mathcal{D}}\triangle\left(\boldsymbol{\mathcal{P}}(\boldsymbol{r})\,\boldsymbol{x}_d(\boldsymbol{r},t)\right) + \boldsymbol{\mathcal{Q}}(\boldsymbol{r})\,\boldsymbol{\mathcal{A}}\boldsymbol{\mathcal{P}}(\boldsymbol{r})\,\boldsymbol{x}_d(\boldsymbol{r},t) + \boldsymbol{\mathcal{Q}}(\boldsymbol{r})\,\boldsymbol{b}.\end{aligned} \tag{5.47}$$

Being a linear partial differential equation for $\boldsymbol{\mathcal{Q}}(\boldsymbol{r})\,\boldsymbol{x}_d(\boldsymbol{r},t)$ with inhomogeneity on the right hand side, Eq. (5.47) can formally be solved with the help of Green's functions. The explicit form of the Green's function depends on the form and dimension of the spatial domain.

Having identified a linearizing assumption, the next step would be to discuss concepts of controllability for linear PDEs, as e.g. the linear diffusion equation, and apply these concepts to Eq. (5.47). However, in contrast to dynamical systems, no condition for controllability in terms of a rank condition for a controllability matrix can be formulated. The reason is that PDEs are essentially dynamical systems with an infinite-dimensional state space. The Cayley-Hamilton theorem cannot be applied to truncate the exponential of a linear operator after a finite number of terms, see Section 2.4.2. We omit a discussion of controllability and present position control of traveling waves as an application of exactly realizable distributions.

5.4. Position control of traveling waves

For simplicity, the projectors $\boldsymbol{\mathcal{Q}}$ and $\boldsymbol{\mathcal{P}}$ are assumed to be constant in space and independent of the state $\boldsymbol{x}$,

$$\boldsymbol{\mathcal{Q}}(\boldsymbol{x},\boldsymbol{r}) = \boldsymbol{\mathcal{Q}} = \text{const.}, \qquad \boldsymbol{\mathcal{P}}(\boldsymbol{x},\boldsymbol{r}) = \boldsymbol{\mathcal{P}} = \text{const.}. \tag{5.48}$$

Note that this implies that the control acts everywhere in position space,

$$\boldsymbol{\chi}(\boldsymbol{r}) = \mathbf{1}. \tag{5.49}$$

Consider the uncontrolled reaction-diffusion system in an unbounded domain $\Omega = \mathbb{R}^N$,

$$\partial_t \boldsymbol{x}(\boldsymbol{r},t) = \boldsymbol{\mathcal{D}}\triangle \boldsymbol{x}(\boldsymbol{r},t) + \boldsymbol{R}\left(\boldsymbol{x}(\boldsymbol{r},t)\right). \tag{5.50}$$

Many reaction-diffusion systems exhibit plane traveling wave solutions propagating with constant velocity c in a constant direction $\hat{\boldsymbol{c}}$, $|\hat{\boldsymbol{c}}| = 1$. A traveling wave is characterized by a wave profile $\boldsymbol{X}_c$ depending only on a single coordinate as

$$\boldsymbol{x}(\boldsymbol{r}, t) = \boldsymbol{X}_c\left(\hat{\boldsymbol{c}}^T\boldsymbol{r} - ct\right) = \boldsymbol{X}_c\left(\sum_{i=1}^{n}\hat{c}_i r_i - ct\right). \tag{5.51}$$

In a frame of reference $\xi = \hat{\boldsymbol{c}}^T\boldsymbol{r} - ct$ comoving with the traveling wave, the wave profile $\boldsymbol{X}_c$ appears stationary and satisfies the profile equation

$$\boldsymbol{0} = \boldsymbol{\mathcal{D}}\boldsymbol{X}_c''(\xi) + c\boldsymbol{X}_c'(\xi) + \boldsymbol{R}(\boldsymbol{X}_c(\xi)). \tag{5.52}$$

The ODE for the wave profile, Eq. (5.52), can exhibit one or more homogeneous steady states. Typically, for $\xi \to \pm\infty$, the wave profile $\boldsymbol{X}_c$ approaches either two different steady states or the same steady state. This fact can be used to classify traveling wave profiles. Front profiles connect different steady states for $\xi \to \pm\infty$ and are found to be heteroclinic orbits of Eq. (5.52), while pulse profiles join the same steady state and are found to be homoclinic orbits. Pulse profiles are naturally localized and usually every component exhibits one or several extrema. Fronts are not localized but typically exhibit a narrow region where the transition from one to the other steady state occurs. Therefore, all traveling wave solutions are localized in the sense that the derivatives of any order $m \geq 1$ of the wave profile $\boldsymbol{X}_c(\xi)$ with respect to the traveling wave coordinate ξ decays to zero,

$$\lim_{\xi\to\pm\infty}\partial_\xi^m\boldsymbol{X}_c(\xi) = \boldsymbol{0}. \tag{5.53}$$

Note that Eq. (5.53) is not a boundary condition for $\boldsymbol{X}_c$ but characterizes the solution to Eq. (5.52).

We assume that before control is switched on at time $t = t_0$, the traveling wave moves unperturbed. Thus, the initial condition for the controlled reaction-diffusion system Eq. (5.9) is

$$\boldsymbol{x}(\boldsymbol{r}, t_0) = \boldsymbol{X}_c\left(\hat{\boldsymbol{c}}^T\boldsymbol{r} - ct_0\right). \tag{5.54}$$

The idea of position control is to choose the desired distribution $\boldsymbol{x}_d(\boldsymbol{r}, t)$ in form of a traveling wave profile $\boldsymbol{X}_c$ shifted according to a protocol of motion $\phi(t)$. The function $\phi(t)$ encodes the desired position over time of the controlled traveling wave along the spatial direction $\hat{\boldsymbol{c}}$. The position of a traveling wave is defined by a distinguishing point of the wave profile. For pulse solutions, the extremum of a certain component of the wave profile defines its position. The position of a front solution is defined by a characteristic point in the transition region as e.g. the point of the steepest slope. A problem arises because for an exactly realizable distribution, only p out of all n components of $\boldsymbol{x}_d(\boldsymbol{r}, t)$ can be prescribed, while the remaining $n-p$ components have to satisfy the constraint equation (5.35). Here, the convention

is that the part $\boldsymbol{\mathcal{P}}\boldsymbol{x}_d(\boldsymbol{r},t)$ is the traveling wave profile $\boldsymbol{X}_c$ shifted according to the protocol of motion $\phi(t)$,

$$\boldsymbol{\mathcal{P}}\boldsymbol{x}_d(\boldsymbol{r},t) = \boldsymbol{\mathcal{P}}\boldsymbol{X}_c\left(\hat{\boldsymbol{c}}^T\boldsymbol{r} - \phi(t)\right). \tag{5.55}$$

The remaining part $\boldsymbol{\mathcal{Q}}\boldsymbol{x}_d(\boldsymbol{r},t)$ has to satisfy the constraint equation (5.35),

$$\partial_t\left(\boldsymbol{\mathcal{Q}}\boldsymbol{x}_d(\boldsymbol{r},t)\right) = \boldsymbol{\mathcal{Q}}\boldsymbol{\mathcal{D}}\triangle\left(\boldsymbol{\mathcal{Q}}\boldsymbol{x}_d(\boldsymbol{r},t)\right) + \boldsymbol{\mathcal{Q}}\boldsymbol{R}\left(\boldsymbol{x}_d(\boldsymbol{r},t)\right) + \boldsymbol{\mathcal{Q}}\boldsymbol{\mathcal{D}}\boldsymbol{\mathcal{P}}\boldsymbol{X}_c''\left(\hat{\boldsymbol{c}}^T\boldsymbol{r} - \phi(t)\right), \tag{5.56}$$

with initial condition

$$\boldsymbol{\mathcal{Q}}\boldsymbol{x}_d(\boldsymbol{r},t_0) = \boldsymbol{\mathcal{Q}}\boldsymbol{X}_c\left(\hat{\boldsymbol{c}}^T\boldsymbol{r} - ct_0\right). \tag{5.57}$$

From the initial condition Eq. (5.54) follows an initial condition for ϕ as

$$\phi(t_0) = ct_0. \tag{5.58}$$

The profile equation (5.52) for $\boldsymbol{X}_c$ is exploited to obtain

$$\begin{aligned}\partial_t\left(\boldsymbol{\mathcal{Q}}\boldsymbol{x}_d(\boldsymbol{r},t)\right) = {}& \boldsymbol{\mathcal{Q}}\boldsymbol{\mathcal{D}}\left(\triangle\boldsymbol{\mathcal{Q}}\boldsymbol{x}_d(\boldsymbol{r},t) - \boldsymbol{\mathcal{Q}}\boldsymbol{X}_c''\left(\hat{\boldsymbol{c}}^T\boldsymbol{r} - \phi(t)\right)\right)\\ &+ \boldsymbol{\mathcal{Q}}\left(\boldsymbol{R}\left(\boldsymbol{x}_d(\boldsymbol{r},t)\right) - \boldsymbol{R}\left(\boldsymbol{X}_c\left(\hat{\boldsymbol{c}}^T\boldsymbol{r} - \phi(t)\right)\right)\right) - c\boldsymbol{\mathcal{Q}}\boldsymbol{X}_c'\left(\hat{\boldsymbol{c}}^T\boldsymbol{r} - \phi(t)\right).\end{aligned} \tag{5.59}$$

Let $\boldsymbol{\mathcal{Q}}\boldsymbol{y}_d(\boldsymbol{r},t)$ be defined by

$$\boldsymbol{x}_d(\boldsymbol{r},t) = \boldsymbol{X}_c\left(\hat{\boldsymbol{c}}^T\boldsymbol{r} - \phi(t)\right) + \boldsymbol{\mathcal{Q}}\boldsymbol{y}_d(\boldsymbol{r},t). \tag{5.60}$$

The constraint equation (5.59) can be written as a PDE for $\boldsymbol{\mathcal{Q}}\boldsymbol{y}_d(\boldsymbol{r},t)$,

$$\begin{aligned}\partial_t\left(\boldsymbol{\mathcal{Q}}\boldsymbol{y}_d(\boldsymbol{r},t)\right) = {}& \boldsymbol{\mathcal{Q}}\boldsymbol{\mathcal{D}}\triangle\boldsymbol{\mathcal{Q}}\boldsymbol{y}_d(\boldsymbol{r},t) + \boldsymbol{\mathcal{Q}}\boldsymbol{R}\left(\boldsymbol{X}_c\left(\hat{\boldsymbol{c}}^T\boldsymbol{r} - \phi(t)\right) + \boldsymbol{\mathcal{Q}}\boldsymbol{y}_d(\boldsymbol{r},t)\right)\\ &+ \left(\dot{\phi}(t) - c\right)\boldsymbol{\mathcal{Q}}\boldsymbol{X}_c'\left(\hat{\boldsymbol{c}}^T\boldsymbol{r} - \phi(t)\right) - \boldsymbol{\mathcal{Q}}\boldsymbol{R}\left(\boldsymbol{X}_c\left(\hat{\boldsymbol{c}}^T\boldsymbol{r} - \phi(t)\right)\right).\end{aligned} \tag{5.61}$$

The next step is the determination of the control signal $\boldsymbol{u}(t)$. Due to $\boldsymbol{\mathcal{B}}^+(\boldsymbol{x})\boldsymbol{\mathcal{P}} = \boldsymbol{\mathcal{B}}^+(\boldsymbol{x})$, $\boldsymbol{\mathcal{P}}$ being constant in time and space, and Eq. (5.60), the control signal Eq. (5.37) can be cast in the form

$$\begin{aligned}\boldsymbol{u}(\boldsymbol{r},t) = {}& -\boldsymbol{\mathcal{B}}^+\left(\boldsymbol{X}_c\left(\hat{\boldsymbol{c}}^T\boldsymbol{r} - \phi(t)\right) + \boldsymbol{\mathcal{Q}}\boldsymbol{y}_d(\boldsymbol{r},t)\right)\left(\boldsymbol{\mathcal{D}}\boldsymbol{X}_c''\left(\hat{\boldsymbol{c}}^T\boldsymbol{r} - \phi(t)\right)\right.\\ &+\boldsymbol{\mathcal{D}}\triangle\boldsymbol{\mathcal{Q}}\boldsymbol{y}_d(\boldsymbol{r},t) + \dot{\phi}(t)\,\boldsymbol{X}_c'\left(\hat{\boldsymbol{c}}^T\boldsymbol{r} - \phi(t)\right)\\ &\left.+\boldsymbol{R}\left(\boldsymbol{X}_c\left(\hat{\boldsymbol{c}}^T\boldsymbol{r} - \phi(t)\right) + \boldsymbol{\mathcal{Q}}\boldsymbol{y}_d(\boldsymbol{r},t)\right)\right).\end{aligned} \tag{5.62}$$

Exploiting again the profile equation (5.52), the last expression becomes

$$\begin{aligned}\boldsymbol{u}(\boldsymbol{r},t) = {}& -\boldsymbol{\mathcal{B}}^+\left(\boldsymbol{X}_c\left(\hat{\boldsymbol{c}}^T\boldsymbol{r} - \phi(t)\right) + \boldsymbol{\mathcal{Q}}\boldsymbol{y}_d(\boldsymbol{r},t)\right)\left(\left(\dot{\phi}(t) - c\right)\boldsymbol{X}_c'\left(\hat{\boldsymbol{c}}^T\boldsymbol{r} - \phi(t)\right)\right)\\ &+\boldsymbol{R}\left(\boldsymbol{X}_c\left(\hat{\boldsymbol{c}}^T\boldsymbol{r} - \phi(t)\right) + \boldsymbol{\mathcal{Q}}\boldsymbol{y}_d(\boldsymbol{r},t)\right) - \boldsymbol{R}\left(\boldsymbol{X}_c\left(\hat{\boldsymbol{c}}^T\boldsymbol{r} - \phi(t)\right)\right)\\ &+\boldsymbol{\mathcal{D}}\triangle\boldsymbol{\mathcal{Q}}\boldsymbol{y}_d(\boldsymbol{r},t)).\end{aligned} \tag{5.63}$$

Equation (5.63) for the control signal together with the constraint equation in the form of Eq. (5.61) is the starting point for the position control of traveling waves in general reaction-diffusion systems. Several special cases leading to simpler expressions can be identified.

An invertible coupling matrix yields $\boldsymbol{\mathcal{Q}} = \mathbf{0}$ and $\boldsymbol{\mathcal{B}}^{+}(\boldsymbol{x}) = \boldsymbol{\mathcal{B}}^{-1}(\boldsymbol{x})$. The constraint equation (5.61) is trivially satisfied, and the control simplifies to

$$\boldsymbol{u}(\boldsymbol{r},t) = \left(c - \dot{\phi}(t)\right) \boldsymbol{\mathcal{B}}^{-1}\left(\boldsymbol{X}_c\left(\hat{\boldsymbol{c}}^T\boldsymbol{r} - \phi(t)\right)\right) \boldsymbol{X}_c'\left(\hat{\boldsymbol{c}}^T\boldsymbol{r} - \phi(t)\right). \tag{5.64}$$

In this case, the control can expressed solely in terms of the traveling wave profile $\boldsymbol{X}_c$ and its velocity c. Any reference to the nonlinearity $\boldsymbol{R}(\boldsymbol{x})$ vanishes. This can be useful if $\boldsymbol{R}(\boldsymbol{x})$ is only approximately known but the wave profile $\boldsymbol{X}_c$ and its velocity c can be measured with sufficient accuracy in experiments. However, the assumption of an equal number of control signals and state components is restrictive and valid only for a limited number of systems. Nevertheless, it is satisfied for all single-component reaction-diffusion systems with a scalar distributed control signal $u(\boldsymbol{r},t)$. As an example, we discuss position control of traveling fronts in the Schlögl model.

Example 5.1: Position control of fronts in the Schlögl model

Consider an autocatalytic chemical reaction mechanism proposed by Schlögl (Schlögl, 1972)

$$A_1 + 2X \underset{k_1^-}{\overset{k_1^+}{\rightleftharpoons}} 3X, \qquad\qquad X \underset{k_2^-}{\overset{k_2^+}{\rightleftharpoons}} A_2. \tag{5.65}$$

Under the assumption that the concentrations $a_{1/2} = \left[A_{1/2}\right]$ of the chemical species $A_{1/2}$ are kept constant in space and time, a nonlinearity $R(x)$ in form of a cubic polynomial

$$\begin{aligned} R(x) &= k_1^+ a_1 x^2 - k_1^- x^3 - k_2^+ x + k_2^- a_2 \\ &= -k(x - x_0)(x - x_1)(x - x_2) \end{aligned} \tag{5.66}$$

dictates the time evolution of the concentration $x = [X]$. For a certain range of parameters, $R(x)$ possesses three real positive roots $0 < x_0 < x_1 < x_2$. In one spatial dimension r, the uncontrolled reaction-diffusion system known as the Schlögl model becomes

$$\partial_t x(r,t) = D\partial_r^2 x(r,t) + R(x(r,t)). \tag{5.67}$$

Although the Schlögl model is introduced in the context of chemical reactions, a scalar reaction-diffusion equation of the form (5.67) with cubic nonlinearity R can be seen as a paradigmatic model for a bistable medium. Such models have

found widespread application far beyond chemical reactions. An important example is the phase field, which is used to model phenomena as diverse as cell motility (Löber et al., 2015), free boundary problems in fluid mechanics (Anderson et al., 1998), and solidification (Boettinger et al., 2002). Initially, Eq. (5.67) has been discussed in 1938 by Zeldovich and Frank-Kamenetsky in connection with flame propagation (Zeldovich and Frank-Kamenetskii, 1938).

The roots x_0, x_1, and x_2 are homogeneous steady states of the system, with the upper (x_2) and lower (x_0) being stable steady states while the root x_1 is unstable. The Schlögl model exhibits a variety of traveling front solutions,

$$x(r,t) = X_c(r - ct), \tag{5.68}$$

propagating with velocity c. The front profile X_c satisfies the profile equation with $\xi = x - ct$

$$0 = DX_c''(\xi) + cX_c'(\xi) + R(X_c(\xi)). \tag{5.69}$$

Front solutions connect the homogeneous steady states as $\lim_{\xi\to\pm\infty}$. A stable traveling front solution connecting the lower and upper stable states x_0 and x_2, respectively, is known analytically and given by

$$X_c(\xi) = \frac{1}{2}(x_0 + x_2) + \frac{1}{2}(x_0 - x_2)\tanh\left(\frac{1}{2\sqrt{2}}\sqrt{\frac{k}{D}}(x_2 - x_0)\xi\right), \tag{5.70}$$

$$c = \sqrt{\frac{Dk}{2}}(x_0 + x_2 - 2x_1). \tag{5.71}$$

Assuming that the concentrations $a_{1/2}$ can be controlled spatio-temporally by the distributed control signal $u(r,t)$ amounts to the substitution

$$a_{1/2} \to a_{1/2} + u(x,t) \tag{5.72}$$

in Eq. (5.67). The controlled reaction-diffusion system is

$$\partial_t x(r,t) = D\partial_r^2 x(r,t) + R(x(r,t)) + B(x(r,t))\,u(r,t). \tag{5.73}$$

Control by a_2 will be additive with constant coupling function $B(x) = k_2^-$, while for control via a_1 the spatio-temporal forcing couples multiplicatively to the RD kinetics and the coupling function $B(x) = k_1^+ x^2$ becomes state dependent. See (Löber et al., 2014) for a discussion of experimental realizations of the controlled Schlögl model.

In the following, we assume control by parameter a_1 such that the coupling function is

$$B(x) = k_1^+ x^2. \tag{5.74}$$

In the context of chemical systems, x is interpreted as a concentration which only attains positive values, $x \geq 0$. As long as $x > 0$, the coupling function as given by Eq. (5.74) is positive and $B(x)$ does not change its rank. Because the Schlögl model is a single component reaction-diffusion system, a single distributed control signal $u(r,t)$ is sufficient to realize any desired distribution which complies with the initial and boundary conditions of the system. With the desired distribution $x_d(r,t)$ given in terms of the traveling wave solution as

$$x_d(r,t) = X_c(r - \phi(t)), \tag{5.75}$$

the solution for the control signal becomes

$$u(r,t) = \left(c - \dot{\phi}(t)\right) \frac{1}{B(X_c(r - \phi(t)))} X_c'(r - \phi(t)). \tag{5.76}$$

The protocol of motion $\phi(t)$ is chosen to move the front back and forth sinusoidally as

$$\phi(t) = A_0 + A\sin(2\pi t/T + A_1). \tag{5.77}$$

The control is applied starting at time $t = t_0$, upon which the front moves unperturbed with velocity c. To achieve a smooth transition of the position ϕ and velocity $\dot{\phi}$ across $t = t_0$, the constants A_0 and A_1 are determined by the conditions

$$\phi(t_0) = \phi_0, \qquad \dot{\phi}(t_0) = c. \tag{5.78}$$

Figure 5.1 shows a snapshot of the controlled front solution (black solid line) and the control signal as given by Eq. (5.76) (red dashed line).

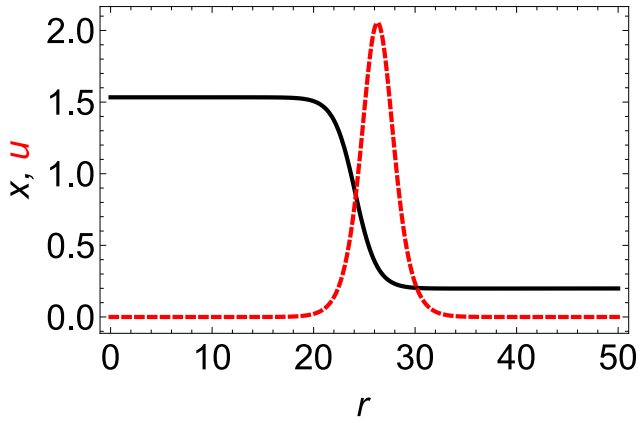

Figure 5.1.: Snapshot of the controlled front solution $x(r,t) = X_c(r - ct)$ to the Schlögl model (black solid line) and distributed control signal (red solid line) as given by Eq. (5.76).

To validate the performance of the control, the protocol of motion $\phi(t)$ is compared with the position over time recorded from numerical simulations of the controlled front. In numerical simulations, the position of the front is defined as the point of the steepest slope of the transition region. Figure 5.2 left demonstrates perfect agreement between prescribed (black solid line) and recorded (red dashed line) position over time. The analogous comparison for the velocity over time shown in Fig. 5.2 right reveals an overall perfect agreement but small deviations at the points of maximum and minimum velocity. Such deviations can be understood to arise from an underlying instability (Löber, 2014). This instability manifests as a finite, non-increasing shift between the positions of control signal and controlled front.

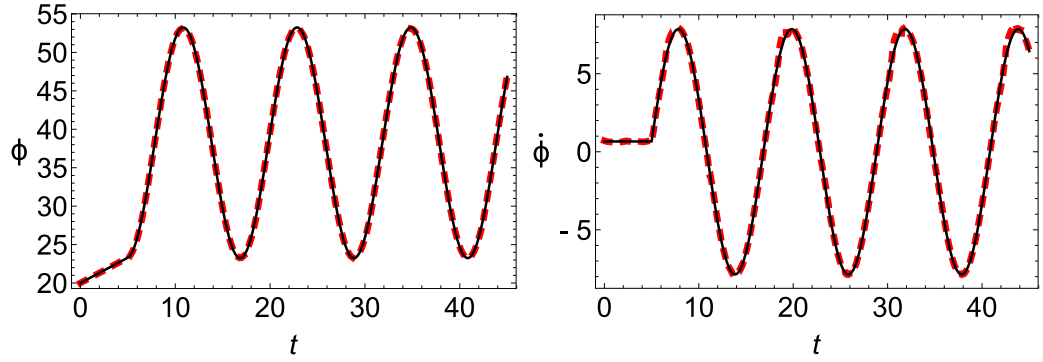

Figure 5.2.: Position control of fronts in the Schlögl model. Left: position over time of the desired protocol of motion (red dashed line) and the actual position over time of the controlled front (black solid line). Right: Velocity over time. Agreement is nearly perfect in both cases.

A second example of position control with a number of control signals smaller than the number of state components is discussed in the following.

Example 5.2: Position control of traveling waves in the activator-controlled FHN model

Consider the one-dimensional spatial domain $0 \leq r < L = 150$ with periodic boundary conditions. Apart from an additional diffusion term, the model equations are the same as in Example 1.2,

$$\begin{pmatrix} \partial_t x(r,t) \\ \partial_t y(r,t) \end{pmatrix} = \begin{pmatrix} D_x \partial_r^2 x(r,t) \\ D_y \partial_r^2 y(r,t) \end{pmatrix} + \begin{pmatrix} a_0 + a_1 x(r,t) + a_2 y(r,t) \\ R(x(r,t), y(r,t)) \end{pmatrix} + \begin{pmatrix} 0 \\ 1 \end{pmatrix} u(r,t). \tag{5.79}$$

The nonlinearity is linear in the inhibitor but nonlinear in the activator, and R is given by

$$R(x,y) = R(y) - x. \tag{5.80}$$

The function $R(y)$ is a cubic polynomial of the form

$$R(y) = 3y - y^3. \tag{5.81}$$

The parameter values are set to

$$a_0 = 0.429, \qquad a_1 = 0, \qquad a_2 = 0.33, \qquad D_y = 1, \qquad D_x = 0.3. \tag{5.82}$$

The traveling wave profile $\boldsymbol{X}_c(\xi) = \begin{pmatrix} X_c(\xi), & Y_c(\xi) \end{pmatrix}^T$ satisfies, with $\xi = x - ct$,

$$D_x X_c''(\xi) + c X_c'(\xi) + a_0 + a_1 X_c(\xi) + a_2 Y_c(\xi) = 0, \tag{5.83}$$

$$D_y Y_c''(\xi) + c Y_c'(\xi) + R(Y_c(\xi)) - X_c(\xi) = 0. \tag{5.84}$$

Exact analytical solutions to Eqs. (5.83) and (5.84) are neither known for finite nor infinite or periodic domains. The wave profile $\boldsymbol{X}_c$ and its velocity c are determined numerically and $\boldsymbol{X}_c$ is interpolated with Mathematica. See Fig. 5.3 for a snapshot of $\boldsymbol{X}_c$.

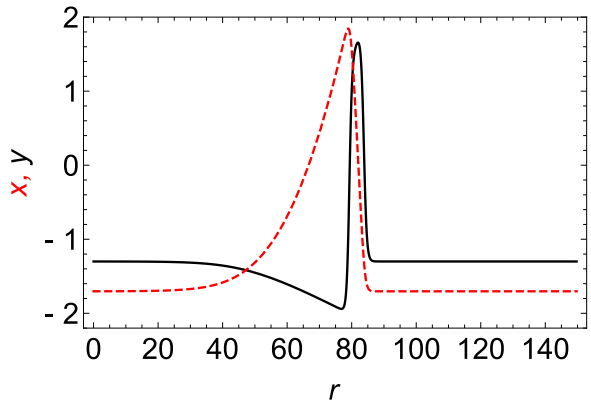

Figure 5.3.: Wave profile of the uncontrolled FHN model moving to the right. Shown is the activator component y (black solid line) and the inhibitor component x (red dashed line). The solution is obtained by numerically solving the uncontrolled reaction-diffusion system, Eq. (5.79) with $u(t) = 0$, for periodic boundary conditions and then interpolated.

The activator component $y_d(r,t)$ of the desired distribution is the traveling wave profile Y_c shifted according to the protocol $\phi(t)$,

$$y_d(r,t) = Y_c(r - \phi(t)), \tag{5.85}$$

while the inhibitor component $x_d(r,t)$ has to satisfy the partial differential equation

$$\partial_t x_d(r,t) - D_x \partial_r^2 x_d(r,t) - a_1 x_d(r,t) = a_0 + a_2 Y_c(r - \phi(t)) \tag{5.86}$$

with initial condition

$$x_d(r,t_0) = X_c(r - \phi(t_0)). \tag{5.87}$$

To simplify Eq. (5.86), $\hat{x}_d$ is defined by the relation

$$x_d(r,t) = \hat{x}_d(r,t) + X_c(r - \phi(t)). \tag{5.88}$$

After using Eq. (5.83), the evolution equation for $\hat{x}_d$ becomes

$$\partial_t \hat{x}_d(r,t) - D_x \partial_r^2 \hat{x}_d(r,t) - a_1 \hat{x}_d(r,t) = -\left(c - \dot{\phi}(t)\right) X_c'(r - \phi(t)), \tag{5.89}$$

$$\hat{x}_d(r,t_0) = 0. \tag{5.90}$$

Together with Eq. (5.84), the control signal is given by the relatively simple expression

$$u(r,t) = \left(c - \dot{\phi}(t)\right) Y_c'(r - \phi(t)) + \hat{x}_d(r,t). \tag{5.91}$$

Exactly the same result as Eq. (5.91) was derived in (Löber and Engel, 2014) with an approach that focused exclusively on position control of traveling waves.

In numerical simulations, the interpolated result for the wave profile is used to formulate the control signal (5.91), and the resulting controlled reaction-diffusion system is solved numerically. Although governed by a linear ODE, the solution for $\hat{x}_d(r,t)$ is determined numerically by solving Eq. (5.89) with periodic boundary conditions for simplicity. The protocol of motion is,

$$\phi(t) = c(t - t_0) + A \sin(2\pi(t - t_0)/T) \tag{5.92}$$

and results in a controlled traveling wave moving sinusoidally back and forth. The values for amplitude and period are $A = 80$ and $T = 20$, respectively. Figure 5.4 compares the desired activator (left) and inhibitor (right, black solid line) with the numerically obtained result of the controlled reaction-diffusion system (red dashed line). On this scale, the agreement is very good. Note that while the controlled activator profile is identical to its uncontrolled profile (see black solid line in Fig. 5.3), the inhibitor wave profile is largely deformed and very different from its uncontrolled counterpart. The reason is simply that only a single state component of the desired distribution can be prescribed, which was chosen to be the activator component, while the inhibitor component is determined by the constraint equation (5.86).

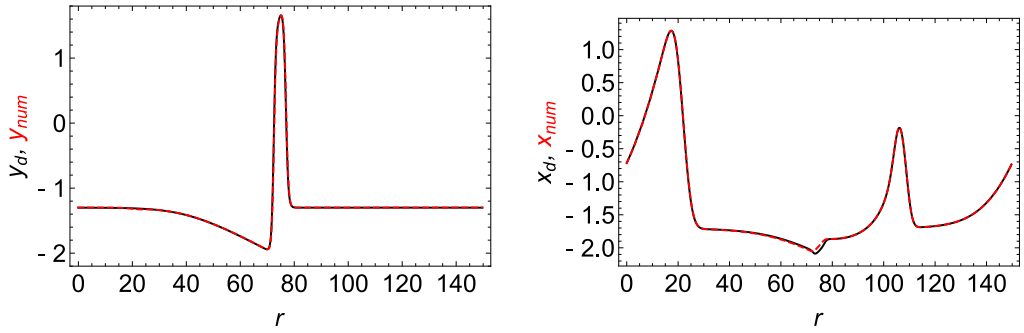

Figure 5.4.: Position control of a traveling pulse in the activator-controlled FHN model. The numerically obtained result for the controlled traveling pulse (black solid line) is very close to the desired distribution (red dashed line). Left: Snapshot of controlled activator y over space r. Right: Snapshot of controlled inhibitor x over space r.

Figure 5.5 shows the difference between the desired and controlled wave profile for activator (left) and inhibitor (right). The differences are *not* in the range of numerical accuracy, and are likely due to an instability. See (Löber, 2014) for a discussion of one possible instability. Because of the steep slopes exhibited by the pulse profile, a small difference in the position between desired and controlled wave has a large effect on the difference between the profiles.

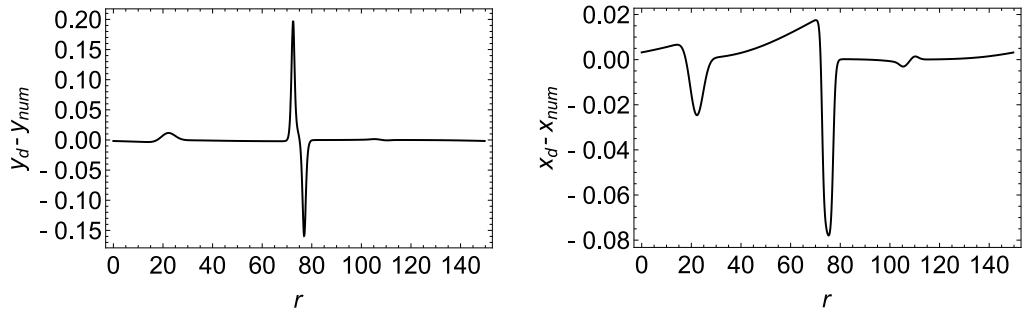

Figure 5.5.: Difference between controlled and desired traveling pulse for the activator (left) and inhibitor (right) component.

Finally, Fig. 5.6 left shows the control signal as given by Eq. (5.91). Being proportional to the derivative of the activator pulse profile Y_c, the control signal has its largest amplitude at the points of the steepest slope of the activator pulse profile. Figure 5.6 right compares the position over time as prescribed by the protocol of motion ϕ (black solid line) with the position over time obtained from numerical simulations (red dashed line). The agreement is well within the range of numerical accuracy. Numerically, the position of the pulse is defined as the position of the maximum of the controlled activator pulse profile.

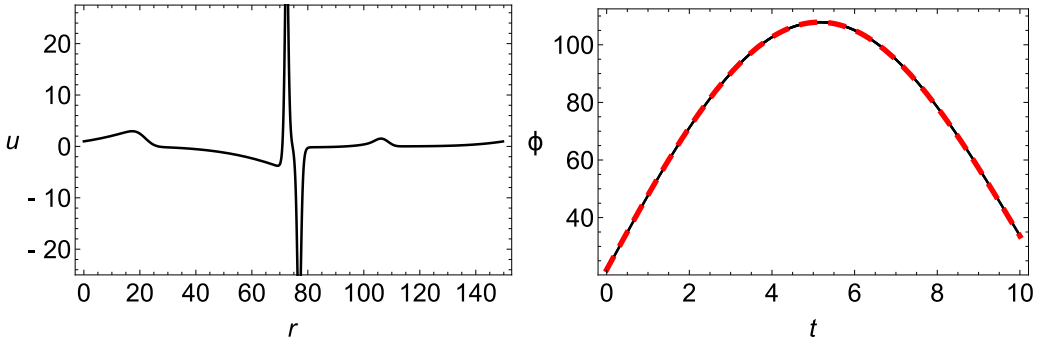

Figure 5.6.: Control signal and controlled position over time for the FHN model. Left: Snapshot of control signal u over space r. Right: Comparison of prescribed protocol ϕ over time (black solid line) and numerically recorded position over time of the controlled pulse (red dashed line).

5.5. Discussion and outlook

5.5.1. Optimal control of reaction-diffusion systems

This section briefly discusses optimal trajectory tracking for reaction-diffusion systems. The mathematical theory of optimal control of PDEs is well developed. The reader is referred to the book (Tröltzsch, 2010) for a mathematically rigorous treatment. Applications of optimal control to reaction-diffusion systems can be found in

(Buchholz et al., 2013; Theißen, 2006; Ryll, 2011).

The target functional for optimal trajectory tracking in reaction-diffusion systems is

$$\begin{aligned}\mathcal{J}\left[\boldsymbol{x}\left(\boldsymbol{r},t\right),\boldsymbol{u}\left(\boldsymbol{r},t\right)\right] =& \frac{1}{2}\int_{t_0}^{t_1} dt \int_{\Omega} d\boldsymbol{r}\left(\boldsymbol{x}\left(\boldsymbol{r},t\right)-\boldsymbol{x}_d\left(\boldsymbol{r},t\right)\right)^T \boldsymbol{\mathcal{S}}\left(\boldsymbol{x}\left(\boldsymbol{r},t\right)-\boldsymbol{x}_d\left(\boldsymbol{r},t\right)\right) \\ &+ \frac{1}{2}\int_{\Omega} d\boldsymbol{r}\left(\boldsymbol{x}\left(\boldsymbol{r},t_1\right)-\boldsymbol{x}_1\left(\boldsymbol{r}\right)\right)^T \boldsymbol{\mathcal{S}}_1\left(\boldsymbol{x}\left(\boldsymbol{r},t_1\right)-\boldsymbol{x}_1\left(\boldsymbol{r}\right)\right) \\ &+ \frac{\epsilon^2}{2}\int_{t_0}^{t_1} dt \int_{\Omega} d\boldsymbol{r}\left(\boldsymbol{u}\left(\boldsymbol{r},t\right)\right)^2. \end{aligned} \tag{5.93}$$

Here, Ω denotes the N-dimensional spatial domain, $\boldsymbol{\mathcal{S}}$ and $\boldsymbol{\mathcal{S}}_1$ are symmetric matrices of weights, and ϵ is the regularization parameter. Apart from the integration over the spatial domain Ω, the functional Eq. (5.93) is identical to the functional Eq. (3.24) for optimal trajectory tracking in dynamical systems from Chapter 3. Equation (5.93) must be minimized under the constraint that $\boldsymbol{x}\left(\boldsymbol{r},t\right)$ is governed by the controlled reaction-diffusion equation

$$\partial_t \boldsymbol{x}\left(\boldsymbol{r},t\right) = \boldsymbol{\mathcal{D}}\triangle\boldsymbol{x}\left(\boldsymbol{r},t\right) + \boldsymbol{R}\left(\boldsymbol{x}\left(\boldsymbol{r},t\right)\right) + \boldsymbol{\mathcal{B}}\left(\boldsymbol{x}\left(\boldsymbol{r},t\right),\boldsymbol{r}\right)\boldsymbol{u}\left(\boldsymbol{r},t\right), \tag{5.94}$$

supplemented with the boundary and initial conditions

$$\boldsymbol{0} = \boldsymbol{n}^T\left(\boldsymbol{r}\right)\left(\boldsymbol{\mathcal{D}}\nabla\boldsymbol{x}\left(\boldsymbol{r},t\right)\right),\ \boldsymbol{r}\in\Gamma, \qquad \boldsymbol{x}\left(\boldsymbol{r},t_0\right) = \boldsymbol{x}_0\left(\boldsymbol{r}\right). \tag{5.95}$$

Similar as in Section 3.1, the constrained minimization problem can be transformed to an unconstrained minimization problem by introducing the vector of Lagrange multipliers $\boldsymbol{\lambda}\left(\boldsymbol{r},t\right)$, also called adjoint state or co-state. This leads to the adjoint or co-state equation for $\boldsymbol{\lambda}$ (Tröltzsch, 2010; Theißen, 2006; Buchholz et al., 2013),

$$\begin{aligned}-\partial_t\boldsymbol{\lambda}\left(\boldsymbol{r},t\right) =& \boldsymbol{\mathcal{D}}\triangle\boldsymbol{\lambda}\left(\boldsymbol{r},t\right) + \left(\nabla\boldsymbol{R}^T\left(\boldsymbol{x}\left(\boldsymbol{r},t\right)\right) + \boldsymbol{u}^T\left(\boldsymbol{r},t\right)\nabla\boldsymbol{\mathcal{B}}^T\left(\boldsymbol{x}\left(\boldsymbol{r},t\right),\boldsymbol{r}\right)\right)\boldsymbol{\lambda}\left(\boldsymbol{r},t\right) \\ &+ \boldsymbol{\mathcal{S}}\left(\boldsymbol{x}\left(\boldsymbol{r},t\right)-\boldsymbol{x}_d\left(\boldsymbol{r},t\right)\right). \end{aligned} \tag{5.96}$$

The co-state $\boldsymbol{\lambda}$ satisfies the same homogeneous Neumann boundary conditions as the state $\boldsymbol{x}$,

$$\boldsymbol{0} = \boldsymbol{n}^T\left(\boldsymbol{r}\right)\left(\boldsymbol{\mathcal{D}}\nabla\boldsymbol{\lambda}\left(\boldsymbol{r},t\right)\right),\ \boldsymbol{r}\in\Gamma, \tag{5.97}$$

and the terminal conditions

$$\boldsymbol{\lambda}\left(\boldsymbol{r},t_1\right) = \boldsymbol{\mathcal{S}}_1\left(\boldsymbol{x}\left(\boldsymbol{r},t_1\right)-\boldsymbol{x}_1\left(\boldsymbol{r}\right)\right). \tag{5.98}$$

Finally, the relation between control signal $\boldsymbol{u}$ and co-state $\boldsymbol{\lambda}$ is obtained as

$$\epsilon^2\boldsymbol{u}\left(\boldsymbol{r},t\right) + \boldsymbol{\mathcal{B}}^T\left(\boldsymbol{x}\left(\boldsymbol{r},t\right),\boldsymbol{r}\right)\boldsymbol{\lambda}\left(\boldsymbol{r},t\right) = \boldsymbol{0}. \tag{5.99}$$

Equations (5.94)-(5.99) constitute the *necessary optimality conditions for optimal trajectory tracking in reaction-diffusion systems.*

For dynamical systems, it was found that the control signal obtained within the framework of exactly realizable trajectories arises as the solution to an unregularized optimal control problem. We expect a similar identity for exactly realizable distributions of spatio-temporal systems. Indeed, if the state equals the desired trajectory everywhere and for all times $t_0 \leq t \leq t_1$, $\boldsymbol{x}(\boldsymbol{r},t) = \boldsymbol{x}_d(\boldsymbol{r},t)$, Eq. (5.96) becomes a homogeneous linear partial differential equation. If additionally the desired distribution complies with the terminal state, $\boldsymbol{x}_d(\boldsymbol{r},t_1) = \boldsymbol{x}_1(\boldsymbol{r})$, the co-state $\boldsymbol{\lambda}(\boldsymbol{r},t)$ vanishes identically everywhere and for all times,

$$\boldsymbol{\lambda}(\boldsymbol{r},t) \equiv \boldsymbol{0}. \tag{5.100}$$

It follows that for a non-vanishing control signal $\boldsymbol{u}(t)$, Eq. (5.99) can only be satisfied if $\epsilon = 0$. In conclusion, all necessary optimality conditions Eqs. (5.94)-(5.99) are satisfied.

However, analogously to the generalized Legendre-Clebsch conditions for dynamical systems, Eqs. (3.83) and (3.84), we expect that there are additional necessary optimality conditions for singular optimal control problems, see Section 3.4.2. While the necessity of the generalized Legendre-Clebsch conditions for dynamical systems is rigorously proven in (Bell and Jacobson, 1975), there seems to be no rigorous proof available for singular optimal control of PDEs. We omit a discussion of additional necessary optimality conditions.

5.5.2. Outlook

A possible next step is the application of the singular perturbation expansion developed in Chapter 4 to the necessary optimality conditions Eqs. (5.94)-(5.99). First, the necessary optimality conditions must be rearranged and split up in equations for the parts $\boldsymbol{\mathcal{P}x}$, $\boldsymbol{\mathcal{Q}x}$, $\boldsymbol{\mathcal{P}\lambda}$, and $\boldsymbol{\mathcal{Q}\lambda}$. Second, the inner and outer equations must be derived. In general, not only time but also space can be rescaled with the small parameter ϵ. This might lead to a larger variety of inner equations and combinations of spatial and temporal boundary layers. However, at least for problems as simple as the activator-controlled FHN model and a control acting everywhere within the spatial domain Ω, it seems reasonable to expect a simple structure of inner and outer equations analogously to the two-dimensional dynamical system from Section 4.1.

The essential difference in the evolution equations between dynamical systems and reaction-diffusion systems is the diffusion term. Being a linear differential operator, we anticipate that if the outer equations of a dynamical system reduce to linear ODEs as the result of a linearizing assumption, the outer equations for a corresponding reaction-diffusion system reduce to linear PDEs. This opens up the interesting possibility to obtain analytical approximations for the optimal control of reaction-diffusion systems. The arising equations will be linear reaction-diffusion equations

with inhomogeneities which involve the desired distribution $\boldsymbol{x}_d$. Such equations can in principle be solved analytically with the help of Green's functions. These solutions would not only provide analytical approximations for open loop control, but would also yield optimal feedback controls for nonlinear reaction-diffusion systems. As was discussed above, it is virtually impossible to numerically deal with optimal feedback control of spatio-temporal systems due to the curse of dimensionality. The approach outlined here would enable an almost exclusive approach to optimal feedback control of nonlinear spatio-temporal systems.

Stability of open loop control methods can never be taken for granted but requires further investigations. The control of exactly realizable distributions might or might not be stable with respect to perturbations of the initial conditions. According to Eq. (5.39) governing the stability of an exactly realizable distribution $\boldsymbol{x}_d(\boldsymbol{r},t)$, the stability of $\boldsymbol{x}_d(\boldsymbol{r},t)$ depends on $\boldsymbol{x}_d(\boldsymbol{r},t)$ itself. This observation opens up the investigation of the stability of position control of traveling waves. Assuming for simplicity a desired distribution of the form

$$\boldsymbol{x}_d(\boldsymbol{r},t) = \boldsymbol{X}_c\left(\hat{\boldsymbol{c}}^T\boldsymbol{r} - \phi(t)\right), \tag{5.101}$$

and a constant coupling matrix $\boldsymbol{\mathcal{B}}(\boldsymbol{x},\boldsymbol{r}) = \boldsymbol{\mathcal{B}} = \text{const.}$, Eq. (5.39), becomes

$$\partial_t \boldsymbol{y}(\boldsymbol{r},t) = \boldsymbol{\mathcal{D}}\triangle\boldsymbol{y}(\boldsymbol{r},t) + \nabla\boldsymbol{R}\left(\boldsymbol{X}_c\left(\hat{\boldsymbol{c}}^T\boldsymbol{r} - \phi(t)\right)\right)\boldsymbol{y}(\boldsymbol{r},t). \tag{5.102}$$

If additionally the protocol velocity is close to the velocity c of the uncontrolled traveling wave, $\dot{\phi}(t) = c + \gamma$ with $|\gamma| \ll 1$, Eq. (5.102) reduces to the equation which determines the linear stability of the traveling wave $\boldsymbol{X}_c$ (Sandstede, 2002). As long as the exactly realizable desired distribution $\boldsymbol{x}_d(\boldsymbol{r},t)$ is sufficiently close to a stable traveling wave solution $\boldsymbol{X}_c$, the controlled wave may be stable. In this way, the controlled wave may benefit from the stability of the uncontrolled traveling wave. A rigorous discussion of stability must take into account the fact that only p out of n components of a desired distribution can be prescribed, and should take into account a state dependent coupling matrix $\boldsymbol{\mathcal{B}}(\boldsymbol{x})$. Additional problems arise because every stable traveling wave possesses at least one eigenvalue with vanishing real part. This fact requires a nonlinear stability analysis. A popular method for that is a multiple scale perturbation expansion (Löber, 2009; Löber et al., 2012). Some aspects of this nonlinear stability analysis for position control of traveling waves are presented in (Löber, 2014), see also the discussion at the end of (Löber et al., 2014). Generally speaking, one can expect a stable open loop control of an exactly realizable desired distribution as long as the desired distribution is sufficiently close to a stable solution of the uncontrolled problem. A thorough understanding of the solutions to an uncontrolled system, including their stability properties, can be very useful for the design of exactly realizable desired distributions which do not require stabilization by additional feedback. Note that the stability analysis of desired distributions which are not exactly realizable is much more difficult. The reason is that the controlled state might be very different from the desired distribution. In general, the stability properties of controlled and desired state are unrelated.

A. Appendix

A.1. General solution for a forced linear dynamical system

Consider a linear n-dimensional dynamical system

$$\dot{\boldsymbol{x}}(t) = \boldsymbol{\mathcal{A}}(t)\,\boldsymbol{x}(t) + \boldsymbol{f}(t), \tag{A.1}$$

and initial conditions

$$\boldsymbol{x}(t_0) = \boldsymbol{x}_0, \tag{A.2}$$

for the state $\boldsymbol{x}$

$$\boldsymbol{x}(t) = (x_1(t), \ldots, x_n(t))^T \tag{A.3}$$

with forcing or inhomogeneity $\boldsymbol{f}$

$$\boldsymbol{f}(t) = (f_1(t), \ldots, f_n(t))^T. \tag{A.4}$$

Dynamical systems of the form Eq. (A.1) are called linear time-variant (LTV) in control theory. If $\boldsymbol{\mathcal{A}}(t) = \boldsymbol{\mathcal{A}} = \text{const.}$ does not depend on time, Eq. (A.1) is called a linear time invariant (LTI) system. See the excellent book (Chen, 1998) and also (Ogata, 2009) for an exhaustive treatment of LTV and LTI systems. The general solution of Eq. (A.1) can be expressed in terms of the principal fundamental $n \times n$ matrix $\boldsymbol{\Phi}(t, t_0)$, also called state transition matrix, which satisfies

$$\partial_t \boldsymbol{\Phi}(t, t_0) = \boldsymbol{\mathcal{A}}(t)\,\boldsymbol{\Phi}(t, t_0), \qquad \boldsymbol{\Phi}(t_0, t_0) = \mathbf{1}. \tag{A.5}$$

$\boldsymbol{\Phi}$ must be a nonsingular matrix such that its inverse $\boldsymbol{\Phi}^{-1}(t_2, t_1) = \boldsymbol{\Phi}(t_1, t_2)$ exists. This implies

$$\boldsymbol{\Phi}^{-1}(t_0, t_0) = \mathbf{1}, \tag{A.6}$$

and

$$\boldsymbol{\Phi}(t_0, t)\,\boldsymbol{\Phi}(t, t_0) = \boldsymbol{\Phi}(t, t_0)\,\boldsymbol{\Phi}(t_0, t) = \mathbf{1}. \tag{A.7}$$

Applying the derivative with respect to time to Eq. (A.7) yields

$$\partial_t \boldsymbol{\Phi}(t_0,t)\,\boldsymbol{\Phi}(t,t_0) = -\boldsymbol{\Phi}(t_0,t)\,\partial_t \boldsymbol{\Phi}(t,t_0) = -\boldsymbol{\Phi}(t_0,t)\,\boldsymbol{\mathcal{A}}(t)\,\boldsymbol{\Phi}(t,t_0)\,. \tag{A.8}$$

From Eq. (A.8) follows the useful relation

$$\boldsymbol{\Phi}(t_0,t)\,\boldsymbol{\mathcal{A}}(t) = -\partial_t \boldsymbol{\Phi}(t_0,t)\,. \tag{A.9}$$

Transposing Eq. (A.9) yields the so-called adjoint equation

$$\partial_t \boldsymbol{\Phi}^T(t_0,t) = -\,\boldsymbol{\mathcal{A}}^T(t)\,\boldsymbol{\Phi}^T(t_0,t)\,. \tag{A.10}$$

Hence, if $\boldsymbol{\Phi}(t,t_0)$ is the fundamental matrix to the original system, then the inverse and transposed matrix $\boldsymbol{\Phi}^{-T}(t,t_0) = \boldsymbol{\Phi}^T(t_0,t)$ is the fundamental matrix to the adjoint system. The general solution $\boldsymbol{x}(t)$ to the inhomogeneous linear system Eq. (A.1) is a superposition

$$\boldsymbol{x}(t) = \boldsymbol{y}(t) + \boldsymbol{z}(t) \tag{A.11}$$

of the solution $\boldsymbol{y}(t)$ to the homogeneous system

$$\dot{\boldsymbol{y}}(t) = \boldsymbol{\mathcal{A}}(t)\,\boldsymbol{y}(t)\,, \qquad\qquad \boldsymbol{y}(t_0) = \boldsymbol{x}_0, \tag{A.12}$$

and a solution $\boldsymbol{z}(t)$ of the inhomogeneous system. The homogeneous solution $\boldsymbol{y}(t)$ can be written in terms of the fundamental matrix as

$$\boldsymbol{y}(t) = \boldsymbol{\Phi}(t,t_0)\,\boldsymbol{x}_0. \tag{A.13}$$

The proof is very simple

$$\dot{\boldsymbol{y}}(t) = \partial_t \boldsymbol{\Phi}(t,t_0)\,\boldsymbol{x}_0 = \boldsymbol{\mathcal{A}}(t)\,\boldsymbol{\Phi}(t,t_0)\,\boldsymbol{x}_0 = \boldsymbol{\mathcal{A}}(t)\,\boldsymbol{y}(t)\,, \tag{A.14}$$

$$\boldsymbol{y}(t_0) = \boldsymbol{\Phi}(t_0,t_0)\,\boldsymbol{x}_0 = \boldsymbol{x}_0. \tag{A.15}$$

The ansatz for the solution of the inhomogeneous linear system Eq. (A.1) is

$$\boldsymbol{z}(t) = \boldsymbol{\Phi}(t,t_0)\,\boldsymbol{v}(t)\,. \tag{A.16}$$

Using the ansatz Eq. (A.16) in the inhomogeneous linear system (A.1) yields

$$\begin{aligned}\dot{\boldsymbol{z}}(t) &= \partial_t \boldsymbol{\Phi}(t,t_0)\,\boldsymbol{v}(t) + \boldsymbol{\Phi}(t,t_0)\,\dot{\boldsymbol{v}}(t) = \boldsymbol{\mathcal{A}}(t)\,\boldsymbol{\Phi}(t,t_0)\,\boldsymbol{v}(t) + \boldsymbol{\Phi}(t,t_0)\,\dot{\boldsymbol{v}}(t) \\ &= \boldsymbol{\mathcal{A}}(t)\,\boldsymbol{\Phi}(t,t_0)\,\boldsymbol{v}(t) + \boldsymbol{f}(t)\,.\end{aligned} \tag{A.17}$$

It follows that

$$\boldsymbol{\Phi}(t,t_0)\,\dot{\boldsymbol{v}}(t) = \boldsymbol{f}(t) \tag{A.18}$$

and, after rearranging and integrating over time,

$$\boldsymbol{v}(t) = \int_{t_0}^{t} d\tau \boldsymbol{\Phi}^{-1}(\tau,t_0)\,\boldsymbol{f}(\tau) = \int_{t_0}^{t} d\tau \boldsymbol{\Phi}(t_0,\tau)\,\boldsymbol{f}(\tau)\,. \tag{A.19}$$

The solution for $\boldsymbol{z}(t)$ is thus

$$\boldsymbol{z}(t) = \boldsymbol{\Phi}(t, t_0)\,\boldsymbol{v}(t) = \boldsymbol{\Phi}(t, t_0) \int_{t_0}^{t} d\tau \boldsymbol{\Phi}(t_0, \tau)\,\boldsymbol{f}(\tau) = \int_{t_0}^{t} d\tau \boldsymbol{\Phi}(t, \tau)\,\boldsymbol{f}(\tau). \quad \text{(A.20)}$$

The general solution $\boldsymbol{x}(t) = \boldsymbol{y}(t) + \boldsymbol{z}(t)$ to Eq. (A.1) is then

$$\boldsymbol{x}(t) = \boldsymbol{\Phi}(t, t_0)\,\boldsymbol{x}_0 + \int_{t_0}^{t} d\tau \boldsymbol{\Phi}(t, \tau)\,\boldsymbol{f}(\tau). \quad \text{(A.21)}$$

For an LTI system with constant state matrix $\boldsymbol{\mathcal{A}}(t) = \boldsymbol{\mathcal{A}} = \text{const.}$, the solution for the state transition matrix $\boldsymbol{\Phi}$ is

$$\boldsymbol{\Phi}(t, t_0) = \exp\left(\boldsymbol{\mathcal{A}}(t - t_0)\right). \quad \text{(A.22)}$$

The matrix exponential is defined by the power series

$$\exp(\boldsymbol{\mathcal{A}}) = \sum_{k=0}^{\infty} \frac{1}{k!} \boldsymbol{\mathcal{A}}^k. \quad \text{(A.23)}$$

A proof of Eq. (A.22) reads as follows. The derivative of $\boldsymbol{\Phi}(t, t_0)$ with respect to time t is

$$\begin{aligned}
\partial_t \boldsymbol{\Phi}(t, t_0) &= \partial_t \exp\left(\boldsymbol{\mathcal{A}}(t - t_0)\right) = \partial_t \left(\sum_{k=0}^{\infty} \frac{1}{k!} \boldsymbol{\mathcal{A}}^k (t - t_0)^k \right) = \sum_{k=0}^{\infty} \frac{1}{k!} \boldsymbol{\mathcal{A}}^k \partial_t (t - t_0)^k \\
&= \sum_{k=1}^{\infty} \frac{k}{k!} \boldsymbol{\mathcal{A}}^k (t - t_0)^{k-1} = \sum_{k=1}^{\infty} \frac{1}{(k-1)!} \boldsymbol{\mathcal{A}}^k (t - t_0)^{k-1} = \sum_{\tilde{k}=0}^{\infty} \frac{1}{\tilde{k}!} \boldsymbol{\mathcal{A}}^{\tilde{k}+1} (t - t_0)^{\tilde{k}} \\
&= \boldsymbol{\mathcal{A}} \sum_{\tilde{k}=0}^{\infty} \frac{1}{\tilde{k}!} \boldsymbol{\mathcal{A}}^{\tilde{k}} (t - t_0)^{\tilde{k}} = \boldsymbol{\mathcal{A}} \exp\left(\boldsymbol{\mathcal{A}}(t - t_0)\right) = \boldsymbol{\mathcal{A}} \boldsymbol{\Phi}(t, t_0). \quad \text{(A.24)}
\end{aligned}$$

The index shift $\tilde{k} = k - 1$ was introduced in the second line.

A.2. Over- and underdetermined systems of linear equations

The solutions of over- and underdetermined systems of linear equations are discussed.

A.2.1. Generalized inverse matrices

The inverse $\boldsymbol{\mathcal{A}}^{-1}$ of a matrix $\boldsymbol{\mathcal{A}}$ with real or complex entries satisfies $\boldsymbol{\mathcal{A}}\boldsymbol{\mathcal{A}}^{-1} = \boldsymbol{\mathcal{A}}^{-1}\boldsymbol{\mathcal{A}} = \mathbf{1}$. An $n \times m$ matrix $\boldsymbol{\mathcal{A}}$ has an inverse only if it is square, i.e., $m = n$, and full rank, i.e., $\mathrm{rank}\,(\boldsymbol{\mathcal{A}}) = n$. For other matrices, a generalized inverse can be defined.

A generalized inverse $\boldsymbol{\mathcal{A}}^g$ of the $n \times m$ matrix $\boldsymbol{\mathcal{A}}$ with real entries has to satisfy the condition

$$\boldsymbol{\mathcal{A}}\boldsymbol{\mathcal{A}}^g\boldsymbol{\mathcal{A}} = \boldsymbol{\mathcal{A}}. \tag{A.25}$$

If $\boldsymbol{\mathcal{A}}^g$ additionally satisfies the condition

$$\boldsymbol{\mathcal{A}}^g\boldsymbol{\mathcal{A}}\boldsymbol{\mathcal{A}}^g = \boldsymbol{\mathcal{A}}^g, \tag{A.26}$$

$\boldsymbol{\mathcal{A}}^g$ is called a generalized reflexive inverse. Furthermore, if $\boldsymbol{\mathcal{A}}^g$ satisfies additionally the conditions

$$(\boldsymbol{\mathcal{A}}\boldsymbol{\mathcal{A}}^g)^T = \boldsymbol{\mathcal{A}}\boldsymbol{\mathcal{A}}^g, \tag{A.27}$$

and

$$(\boldsymbol{\mathcal{A}}^g\boldsymbol{\mathcal{A}})^T = \boldsymbol{\mathcal{A}}^g\boldsymbol{\mathcal{A}}, \tag{A.28}$$

$\boldsymbol{\mathcal{A}}^g$ is called the Moore-Penrose pseudo inverse matrix and denoted by $\boldsymbol{\mathcal{A}}^+$. For any matrix $\boldsymbol{\mathcal{A}}$ with real or complex entries, the Moore-Penrose pseudo inverse $\boldsymbol{\mathcal{A}}^+$ exists and is unique. A generalized inverse satisfying only condition (A.25) is usually not unique (Campbell and Meyer Jr., 1991).

A.2.2. Solving an overdetermined system of linear equations

An overdetermined system of equations has more equations than unknowns. Let $\boldsymbol{x} \in \mathbb{R}^p$ and $\boldsymbol{b} \in \mathbb{R}^n$ with $p < n$, and let $\boldsymbol{\mathcal{A}}$ be an $n \times p$ matrix. The aim is to solve the system of n equations

$$\boldsymbol{\mathcal{A}}\boldsymbol{x} = \boldsymbol{b} \tag{A.29}$$

for $\boldsymbol{x}$. Such overdetermined equations regularly occur in data fitting problems. Because $\boldsymbol{\mathcal{A}}$ is not a quadratic matrix, an exact solution cannot exist. However, a useful expression for $\boldsymbol{x}$ can be derived as follows. Multiplying Eq. (A.29) with $\boldsymbol{\mathcal{A}}^T$ yields

$$\boldsymbol{\mathcal{A}}^T\boldsymbol{\mathcal{A}}\boldsymbol{x} = \boldsymbol{\mathcal{A}}^T\boldsymbol{b}. \tag{A.30}$$

To solve for $\boldsymbol{x}$, Eq. (A.30) is multiplied with the inverse of the $p \times p$ matrix $\boldsymbol{\mathcal{A}}^T\boldsymbol{\mathcal{A}}$ from the left to get

$$\boldsymbol{x} = \left(\boldsymbol{\mathcal{A}}^T\boldsymbol{\mathcal{A}}\right)^{-1}\boldsymbol{\mathcal{A}}^T\boldsymbol{b} = \boldsymbol{\mathcal{A}}^+\boldsymbol{b}. \tag{A.31}$$

The $p \times n$ matrix $\boldsymbol{\mathcal{A}}^+$ is defined as

$$\boldsymbol{\mathcal{A}}^+ = \left(\boldsymbol{\mathcal{A}}^T\boldsymbol{\mathcal{A}}\right)^{-1}\boldsymbol{\mathcal{A}}^T. \tag{A.32}$$

The matrix $\boldsymbol{\mathcal{A}}^+$ is the Moore-Penrose pseudo inverse of matrix $\boldsymbol{\mathcal{A}}$, which can be proven by checking all four conditions Eqs. (A.25)-(A.28). The inverse of $\boldsymbol{\mathcal{A}}^T\boldsymbol{\mathcal{A}}$ exists whenever $\boldsymbol{\mathcal{A}}$ has full column rank p,

$$\operatorname{rank}(\boldsymbol{\mathcal{A}}) = p. \tag{A.33}$$

If $p = n$ and $\boldsymbol{\mathcal{A}}$ has full rank, the inverse of $\boldsymbol{\mathcal{A}}$ exists and

$$\boldsymbol{\mathcal{A}}^+ = \left(\boldsymbol{\mathcal{A}}^T\boldsymbol{\mathcal{A}}\right)^{-1}\boldsymbol{\mathcal{A}}^T = \boldsymbol{\mathcal{A}}^{-1}\boldsymbol{\mathcal{A}}^{-T}\boldsymbol{\mathcal{A}}^T = \boldsymbol{\mathcal{A}}^{-1}. \tag{A.34}$$

Multiplying the expression (A.31) for $\boldsymbol{x}$ from the left by $\boldsymbol{\mathcal{A}}$ as on the l. h. s. of Eq. (A.29) gives

$$\boldsymbol{\mathcal{A}}\boldsymbol{x} = \boldsymbol{\mathcal{A}}\boldsymbol{\mathcal{A}}^+\boldsymbol{b} = \boldsymbol{\mathcal{A}}\left(\boldsymbol{\mathcal{A}}^T\boldsymbol{\mathcal{A}}\right)^{-1}\boldsymbol{\mathcal{A}}^T\boldsymbol{b}. \tag{A.35}$$

Note that

$$\boldsymbol{\mathcal{P}} = \boldsymbol{\mathcal{A}}\boldsymbol{\mathcal{A}}^+ = \boldsymbol{\mathcal{A}}\left(\boldsymbol{\mathcal{A}}^T\boldsymbol{\mathcal{A}}\right)^{-1}\boldsymbol{\mathcal{A}}^T \tag{A.36}$$

is a projector, i. e., it is an idempotent $n \times n$ matrix,

$$\boldsymbol{\mathcal{P}}^2 = \boldsymbol{\mathcal{A}}\left(\boldsymbol{\mathcal{A}}^T\boldsymbol{\mathcal{A}}\right)^{-1}\boldsymbol{\mathcal{A}}^T\boldsymbol{\mathcal{A}}\left(\boldsymbol{\mathcal{A}}^T\boldsymbol{\mathcal{A}}\right)^{-1}\boldsymbol{\mathcal{A}}^T = \boldsymbol{\mathcal{A}}\left(\boldsymbol{\mathcal{A}}^T\boldsymbol{\mathcal{A}}\right)^{-1}\boldsymbol{\mathcal{A}}^T = \boldsymbol{\mathcal{P}}. \tag{A.37}$$

Furthermore, $\boldsymbol{\mathcal{P}}$ is symmetric

$$\boldsymbol{\mathcal{P}}^T = \left(\boldsymbol{\mathcal{A}}\left(\boldsymbol{\mathcal{A}}^T\boldsymbol{\mathcal{A}}\right)^{-1}\boldsymbol{\mathcal{A}}^T\right)^T = \boldsymbol{\mathcal{A}}\left(\boldsymbol{\mathcal{A}}^T\boldsymbol{\mathcal{A}}\right)^{-T}\boldsymbol{\mathcal{A}}^T = \boldsymbol{\mathcal{A}}\left(\boldsymbol{\mathcal{A}}^T\boldsymbol{\mathcal{A}}\right)^{-1}\boldsymbol{\mathcal{A}}^T = \boldsymbol{\mathcal{P}}. \tag{A.38}$$

Note that the inverse of the symmetric matrix $\boldsymbol{\mathcal{A}}^T\boldsymbol{\mathcal{A}}$ is also symmetric. The projector $\boldsymbol{\mathcal{P}}$ has rank

$$\operatorname{rank}(\boldsymbol{\mathcal{P}}) = p. \tag{A.39}$$

A projector $\boldsymbol{\mathcal{Q}}$ complementary to $\boldsymbol{\mathcal{P}}$ can be defined as

$$\boldsymbol{\mathcal{Q}} = \mathbf{1} - \boldsymbol{\mathcal{P}}, \tag{A.40}$$

which is also idempotent and symmetric and has rank

$$\operatorname{rank}(\boldsymbol{\mathcal{Q}}) = n - p. \tag{A.41}$$

With the help of these projectors, the l. h. s. of Eq. (A.29) can be written as

$$\boldsymbol{\mathcal{A}}\boldsymbol{x} = \boldsymbol{\mathcal{P}}\boldsymbol{b} = \boldsymbol{b} - \boldsymbol{\mathcal{Q}}\boldsymbol{b}. \tag{A.42}$$

According to (A.29), this should be equal to $\boldsymbol{b}$, which, of course, can only be true if

$$\boldsymbol{\mathcal{Q}}\boldsymbol{b} = \boldsymbol{0}. \tag{A.43}$$

In general, Eq. (A.43) is not true, and therefore the "solution" Eq. (A.31) cannot be an exact solution. In fact, Eq. (A.43) is the condition for an exact solution to exist. That means that either $\boldsymbol{b}$ is the null vector, or the matrix $\boldsymbol{\mathcal{Q}}$ is the null matrix. The third possibility is that $\boldsymbol{b}$ lies in the null space of $\boldsymbol{\mathcal{Q}}$.

The expression Eq. (A.31) can be understood to give an optimal approximate solution in the least square sense. In the following, we demonstrate that $\boldsymbol{x} = \boldsymbol{\mathcal{A}}^{+}\boldsymbol{b}$ is the solution to the minimization problem

$$\min_{\boldsymbol{x}} \frac{1}{2}\left(\boldsymbol{\mathcal{A}}\boldsymbol{x} - \boldsymbol{b}\right)^2. \tag{A.44}$$

Define the scalar function $\mathcal{J}$ as

$$\mathcal{J}(\boldsymbol{x}) = \frac{1}{2}\left(\boldsymbol{\mathcal{A}}\boldsymbol{x} - \boldsymbol{b}\right)^2 = \frac{1}{2}\left(\boldsymbol{x}^T\boldsymbol{\mathcal{A}}^T\boldsymbol{\mathcal{A}}\boldsymbol{x} - 2\boldsymbol{b}^T\boldsymbol{\mathcal{A}}\boldsymbol{x} + \boldsymbol{b}^T\boldsymbol{b}\right). \tag{A.45}$$

The Jacobian $\nabla\mathcal{J}$ of $\mathcal{J}$ with respect to $\boldsymbol{x}$ is given by

$$\nabla\mathcal{J}(\boldsymbol{x}) = \boldsymbol{x}^T\boldsymbol{\mathcal{A}}^T\boldsymbol{\mathcal{A}} - \boldsymbol{b}^T\boldsymbol{\mathcal{A}}. \tag{A.46}$$

The function $\mathcal{J}$ attains its extremum whenever

$$\nabla\mathcal{J}(\boldsymbol{x}) = \boldsymbol{0}. \tag{A.47}$$

Consequently, the vector $\boldsymbol{x}$ for which $\mathcal{J}$ attains its extremum must satisfy the equation

$$\boldsymbol{x}^T\boldsymbol{\mathcal{A}}^T\boldsymbol{\mathcal{A}} = \boldsymbol{b}^T\boldsymbol{\mathcal{A}}, \tag{A.48}$$

or, after transposing,

$$\boldsymbol{\mathcal{A}}^T\boldsymbol{\mathcal{A}}\boldsymbol{x} = \boldsymbol{\mathcal{A}}^T\boldsymbol{b}. \tag{A.49}$$

Solving for $\boldsymbol{x}$ indeed yields the expression Eq. (A.31),

$$\boldsymbol{x} = \left(\boldsymbol{\mathcal{A}}^T\boldsymbol{\mathcal{A}}\right)^{-1}\boldsymbol{\mathcal{A}}^T\boldsymbol{b} = \boldsymbol{\mathcal{A}}^{+}\boldsymbol{b}. \tag{A.50}$$

It remains to check if the extremum is indeed a minimum. Computing the Hessian matrix $\nabla^2 \mathcal{J}$ of $\mathcal{J}$ yields

$$\nabla^2 \mathcal{J}(\boldsymbol{x}) = \boldsymbol{\mathcal{A}}^T \boldsymbol{\mathcal{A}}. \tag{A.51}$$

For an arbitrary matrix $\boldsymbol{\mathcal{A}}$, $\boldsymbol{\mathcal{A}}^T \boldsymbol{\mathcal{A}}$ is a positive semidefinite matrix. It becomes a positive definite matrix if $\boldsymbol{\mathcal{A}}^T \boldsymbol{\mathcal{A}}$ is nonsingular, or, equivalently, if $\boldsymbol{\mathcal{A}}$ has full rank, $\operatorname{rank}(\boldsymbol{\mathcal{A}}) = p$ (Chen, 1998). Therefore, $\boldsymbol{x} = \boldsymbol{\mathcal{A}}^+ \boldsymbol{b}$ indeed minimizes $\mathcal{J}$.

In conclusion, the linear equation $\boldsymbol{\mathcal{A}}\boldsymbol{x} = \boldsymbol{b}$ is discussed. An optimal solution for $\boldsymbol{x}$, which minimizes the squared difference $(\boldsymbol{\mathcal{A}}\boldsymbol{x} - \boldsymbol{b})^2$, exists as long as $\boldsymbol{\mathcal{A}}^T \boldsymbol{\mathcal{A}}$ is positive definite and is given by $\boldsymbol{x} = \left(\boldsymbol{\mathcal{A}}^T \boldsymbol{\mathcal{A}}\right)^{-1} \boldsymbol{\mathcal{A}}^T \boldsymbol{b}$. An exact solution for $\boldsymbol{x}$ can only exist if additionally, the vector $\boldsymbol{b}$ satisfies the constraint $\left(\mathbf{1} - \boldsymbol{\mathcal{A}} \left(\boldsymbol{\mathcal{A}}^T \boldsymbol{\mathcal{A}}\right)^{-1} \boldsymbol{\mathcal{A}}^T\right) \boldsymbol{b} = \mathbf{0}$.

Finally, a slightly more general minimization problem is discussed. The problem is to minimize

$$\min_{\boldsymbol{x}} \frac{1}{2} (\boldsymbol{\mathcal{A}}\boldsymbol{x} - \boldsymbol{b})^T \boldsymbol{\mathcal{S}} (\boldsymbol{\mathcal{A}}\boldsymbol{x} - \boldsymbol{b}), \tag{A.52}$$

with the symmetric $n \times n$ matrix $\boldsymbol{\mathcal{S}}^T = \boldsymbol{\mathcal{S}}$ of weighting coefficients. In the same manner as before, the scalar function $\mathcal{J}_{\boldsymbol{\mathcal{S}}}$ is defined as

$$\mathcal{J}_{\boldsymbol{\mathcal{S}}}(\boldsymbol{x}) = \frac{1}{2} (\boldsymbol{\mathcal{A}}\boldsymbol{x} - \boldsymbol{b})^T \boldsymbol{\mathcal{S}} (\boldsymbol{\mathcal{A}}\boldsymbol{x} - \boldsymbol{b}) = \frac{1}{2} \left(\boldsymbol{x}^T \boldsymbol{\mathcal{A}}^T \boldsymbol{\mathcal{S}} \boldsymbol{\mathcal{A}} \boldsymbol{x} - 2\boldsymbol{b}^T \boldsymbol{\mathcal{S}} \boldsymbol{\mathcal{A}} \boldsymbol{x} + \boldsymbol{b}^T \boldsymbol{\mathcal{S}} \boldsymbol{b}\right). \tag{A.53}$$

The function $\mathcal{J}_{\boldsymbol{\mathcal{S}}}$ attains its extremum if

$$\nabla \mathcal{J}_{\boldsymbol{\mathcal{S}}}(\boldsymbol{x}) = \mathbf{0}, \tag{A.54}$$

which gives

$$\boldsymbol{\mathcal{A}}^T \boldsymbol{\mathcal{S}} \boldsymbol{\mathcal{A}} \boldsymbol{x} = \boldsymbol{\mathcal{A}}^T \boldsymbol{\mathcal{S}} \boldsymbol{b}. \tag{A.55}$$

As long as the $p \times p$ matrix $\boldsymbol{\mathcal{A}}^T \boldsymbol{\mathcal{S}} \boldsymbol{\mathcal{A}}$ has full rank,

$$\operatorname{rank}\left(\boldsymbol{\mathcal{A}}^T \boldsymbol{\mathcal{S}} \boldsymbol{\mathcal{A}}\right) = p, \tag{A.56}$$

Eq. (A.55) can be solved for $\boldsymbol{x}$ to get

$$\boldsymbol{x} = \left(\boldsymbol{\mathcal{A}}^T \boldsymbol{\mathcal{S}} \boldsymbol{\mathcal{A}}\right)^{-1} \boldsymbol{\mathcal{A}}^T \boldsymbol{\mathcal{S}} \boldsymbol{b} = \boldsymbol{\mathcal{A}}_{\boldsymbol{\mathcal{S}}}^+ \boldsymbol{b}. \tag{A.57}$$

The generalized inverse $p \times n$ matrix $\boldsymbol{\mathcal{A}}_{\boldsymbol{\mathcal{S}}}^+$ is defined as

$$\boldsymbol{\mathcal{A}}_{\boldsymbol{\mathcal{S}}}^+ = \left(\boldsymbol{\mathcal{A}}^T \boldsymbol{\mathcal{S}} \boldsymbol{\mathcal{A}}\right)^{-1} \boldsymbol{\mathcal{A}}^T \boldsymbol{\mathcal{S}}. \tag{A.58}$$

The question arises if $\boldsymbol{\mathcal{A}}_{\boldsymbol{\mathcal{S}}}^{+}$ is the Moore-Penrose pseudo inverse. Checking the four conditions Eqs. (A.25)-(A.28) reveals that all conditions except Eq. (A.27) are satisfied. Consequently, $\boldsymbol{\mathcal{A}}_{\boldsymbol{\mathcal{S}}}^{+}$ is not a Moore-Penrose pseudo inverse but a generalized reflexive inverse. What remains to check is if the extremum is indeed a minimum. Computing the Hessian matrix $\nabla\nabla\mathcal{J}_{\boldsymbol{\mathcal{S}}}$ of $\mathcal{J}_{\boldsymbol{\mathcal{S}}}$ yields

$$\nabla\nabla\mathcal{J}_{\boldsymbol{\mathcal{S}}}\left(\boldsymbol{x}\right) = \boldsymbol{\mathcal{A}}^{T}\boldsymbol{\mathcal{S}}\boldsymbol{\mathcal{A}}. \tag{A.59}$$

Consequently, as long as $\boldsymbol{\mathcal{A}}^{T}\boldsymbol{\mathcal{S}}\boldsymbol{\mathcal{A}}$ is positive definite, the solution Eq. (A.57) minimizes $\mathcal{J}_{\boldsymbol{\mathcal{S}}}$. Note that a positive definite matrix has always full rank and is invertible, such that the solution Eq. (A.57) exists. Similar as above, two complementary projectors $\boldsymbol{\mathcal{P}}_{\boldsymbol{\mathcal{S}}}$ and $\boldsymbol{\mathcal{Q}}_{\boldsymbol{\mathcal{S}}}$ can be defined as

$$\boldsymbol{\mathcal{P}}_{\boldsymbol{\mathcal{S}}} = \boldsymbol{\mathcal{A}}\boldsymbol{\mathcal{A}}_{\boldsymbol{\mathcal{S}}}^{+} = \boldsymbol{\mathcal{A}}\left(\boldsymbol{\mathcal{A}}^{T}\boldsymbol{\mathcal{S}}\boldsymbol{\mathcal{A}}\right)^{-1}\boldsymbol{\mathcal{A}}^{T}\boldsymbol{\mathcal{S}}, \qquad \boldsymbol{\mathcal{Q}}_{\boldsymbol{\mathcal{S}}} = \mathbf{1} - \boldsymbol{\mathcal{P}}_{\boldsymbol{\mathcal{S}}}. \tag{A.60}$$

In contrast to the projectors $\boldsymbol{\mathcal{P}}$ and $\boldsymbol{\mathcal{Q}}$, these projectors are not symmetric. For the optimal solution $\boldsymbol{x}$ to be an exact solution to $\boldsymbol{\mathcal{A}}\boldsymbol{x} = \boldsymbol{b}$, the vector $\boldsymbol{b}$ has to satisfy an additional condition,

$$\boldsymbol{\mathcal{Q}}_{\boldsymbol{\mathcal{S}}}\boldsymbol{b} = \mathbf{0}. \tag{A.61}$$

Thus, two exact solutions to the linear equation $\boldsymbol{\mathcal{A}}\boldsymbol{x} = \boldsymbol{b}$ were found. The first solution is given by

$$\boldsymbol{x} = \boldsymbol{x}_1 = \left(\boldsymbol{\mathcal{A}}^{T}\boldsymbol{\mathcal{A}}\right)^{-1}\boldsymbol{\mathcal{A}}^{T}\boldsymbol{b}, \qquad \boldsymbol{\mathcal{Q}}\boldsymbol{b} = \mathbf{0}, \tag{A.62}$$

while the second solution is

$$\boldsymbol{x} = \boldsymbol{x}_2 = \left(\boldsymbol{\mathcal{A}}^{T}\boldsymbol{\mathcal{S}}\boldsymbol{\mathcal{A}}\right)^{-1}\boldsymbol{\mathcal{A}}^{T}\boldsymbol{\mathcal{S}}\boldsymbol{b}, \qquad \boldsymbol{\mathcal{Q}}_{\boldsymbol{\mathcal{S}}}\boldsymbol{b} = \mathbf{0}. \tag{A.63}$$

The exact solution to $\boldsymbol{\mathcal{A}}\boldsymbol{x} = \boldsymbol{b}$ should be unique such that

$$\boldsymbol{x}_1 = \boldsymbol{x}_2. \tag{A.64}$$

Indeed, computing their difference, multiplying by $\boldsymbol{\mathcal{A}}$, and exploiting the relations for $\boldsymbol{b}$ yields

$$\begin{aligned}\boldsymbol{\mathcal{A}}\left(\boldsymbol{x}_1 - \boldsymbol{x}_2\right) &= \boldsymbol{\mathcal{A}}\left(\boldsymbol{\mathcal{A}}^{T}\boldsymbol{\mathcal{A}}\right)^{-1}\boldsymbol{\mathcal{A}}^{T}\boldsymbol{b} - \boldsymbol{\mathcal{A}}\left(\boldsymbol{\mathcal{A}}^{T}\boldsymbol{\mathcal{S}}\boldsymbol{\mathcal{A}}\right)^{-1}\boldsymbol{\mathcal{A}}^{T}\boldsymbol{\mathcal{S}}\boldsymbol{b} \\ &= \left(\boldsymbol{\mathcal{P}} - \boldsymbol{\mathcal{P}}_{\boldsymbol{\mathcal{S}}}\right)\boldsymbol{b} = \left(\mathbf{1} - \boldsymbol{\mathcal{Q}} - \mathbf{1} + \boldsymbol{\mathcal{Q}}_{\boldsymbol{\mathcal{S}}}\right)\boldsymbol{b} = \mathbf{0}.\end{aligned} \tag{A.65}$$

This relation is true if either $\boldsymbol{x}_1 = \boldsymbol{x}_2$, or $\boldsymbol{x}_1 - \boldsymbol{x}_2$ lies in the null space of $\boldsymbol{\mathcal{A}}$. However, due to the assumption that $\boldsymbol{\mathcal{A}}$ has full rank and as a consequence of the rank-nullity theorem, the null space of $\boldsymbol{\mathcal{A}}$ has zero dimension and contains only the zero vector. Consequently, the solutions are identical, $\boldsymbol{x}_1 = \boldsymbol{x}_2$. As expected, the exact solution is unique, and does not depend on the matrix of weighting coefficients $\boldsymbol{\mathcal{S}}$. The relations $\boldsymbol{\mathcal{Q}}\boldsymbol{b} = \mathbf{0}$ and $\boldsymbol{\mathcal{Q}}_{\boldsymbol{\mathcal{S}}}\boldsymbol{b} = \mathbf{0}$ are the analogues of the constraint equations for exactly realizable desired trajectories introduced in Section 2.2 and Section 3.4, respectively.

A.2.3. Solving an underdetermined system of equations

The opposite problem is considered. An underdetermined system is a system with fewer equations than unknowns. Let $\boldsymbol{x} \in \mathbb{R}^p$ and $\boldsymbol{b} \in \mathbb{R}^n$ with $p > n$, and let $\boldsymbol{\mathcal{A}}$ be an $n \times p$ matrix. The system of n equations

$$\boldsymbol{\mathcal{A}}\boldsymbol{x} = \boldsymbol{b} \tag{A.66}$$

is to be solved for $\boldsymbol{x}$. Because there are fewer equations than components of $\boldsymbol{x}$, not all components of $\boldsymbol{x}$ can be determined. Motivated by the example above, two complementary projectors $\boldsymbol{\mathcal{M}}$ and $\boldsymbol{\mathcal{N}}$ are introduced as

$$\boldsymbol{\mathcal{M}} = \boldsymbol{\mathcal{A}}^T \left(\boldsymbol{\mathcal{A}}\boldsymbol{\mathcal{A}}^T\right)^{-1} \boldsymbol{\mathcal{A}}, \tag{A.67}$$

$$\boldsymbol{\mathcal{N}} = \mathbf{1} - \boldsymbol{\mathcal{M}}. \tag{A.68}$$

These projectors are symmetric $p \times p$ matrices. Note that the $n \times n$ matrix $\boldsymbol{\mathcal{A}}\boldsymbol{\mathcal{A}}^T$ has rank n whenever $\boldsymbol{\mathcal{A}}$ has full row rank,

$$\operatorname{rank}(\boldsymbol{\mathcal{A}}) = n. \tag{A.69}$$

The projectors $\boldsymbol{\mathcal{M}}$ and $\boldsymbol{\mathcal{N}}$ have rank

$$\operatorname{rank}(\boldsymbol{\mathcal{M}}) = n, \qquad \operatorname{rank}(\boldsymbol{\mathcal{N}}) = p - n. \tag{A.70}$$

Multiplying Eq. (A.66) by $\boldsymbol{\mathcal{A}}^T \left(\boldsymbol{\mathcal{A}}\boldsymbol{\mathcal{A}}^T\right)^{-1}$ from the left yields

$$\boldsymbol{\mathcal{A}}^T \left(\boldsymbol{\mathcal{A}}\boldsymbol{\mathcal{A}}^T\right)^{-1} \boldsymbol{b} = \boldsymbol{\mathcal{A}}^T \left(\boldsymbol{\mathcal{A}}\boldsymbol{\mathcal{A}}^T\right)^{-1} \boldsymbol{\mathcal{A}}\boldsymbol{x} = \boldsymbol{\mathcal{M}}\boldsymbol{x}. \tag{A.71}$$

Thus, the part $\boldsymbol{\mathcal{M}}\boldsymbol{x}$ can be determined in terms of $\boldsymbol{b}$, while the part $\boldsymbol{\mathcal{N}}\boldsymbol{x}$ must be left undetermined.

A.3. Properties of time-dependent projectors

Some relations for the projectors $\boldsymbol{\mathcal{P}}(\boldsymbol{x})$ and $\boldsymbol{\mathcal{Q}}(\boldsymbol{x})$ are listed. The projectors may depend on time though its argument $\boldsymbol{x}$. First, the projectors are idempotent,

$$\boldsymbol{\mathcal{Q}}(\boldsymbol{x})\,\boldsymbol{\mathcal{Q}}(\boldsymbol{x}) = \boldsymbol{\mathcal{Q}}(\boldsymbol{x}), \qquad \boldsymbol{\mathcal{P}}(\boldsymbol{x})\,\boldsymbol{\mathcal{P}}(\boldsymbol{x}) = \boldsymbol{\mathcal{P}}(\boldsymbol{x}), \tag{A.72}$$

and complementary,

$$\boldsymbol{\mathcal{P}}(\boldsymbol{x}) + \boldsymbol{\mathcal{Q}}(\boldsymbol{x}) = \mathbf{1}. \tag{A.73}$$

Applying the time derivative $\frac{d}{dt}$ to the last relation yields

$$\frac{d}{dt}\boldsymbol{\mathcal{P}}\left(\boldsymbol{x}\left(t\right)\right) = -\frac{d}{dt}\boldsymbol{\mathcal{Q}}\left(\boldsymbol{x}\left(t\right)\right) \tag{A.74}$$

or

$$\nabla\boldsymbol{\mathcal{P}}\left(\boldsymbol{x}\left(t\right)\right)\dot{\boldsymbol{x}}\left(t\right) = -\nabla\boldsymbol{\mathcal{Q}}\left(\boldsymbol{x}\left(t\right)\right)\dot{\boldsymbol{x}}\left(t\right) \tag{A.75}$$

or

$$\nabla\boldsymbol{\mathcal{P}}\left(\boldsymbol{x}\left(t\right)\right) = -\nabla\boldsymbol{\mathcal{Q}}\left(\boldsymbol{x}\left(t\right)\right). \tag{A.76}$$

Here, $\nabla\boldsymbol{\mathcal{P}}\left(\boldsymbol{x}\right)$ denotes the Jacobian of $\boldsymbol{\mathcal{P}}\left(\boldsymbol{x}\right)$ with respect to $\boldsymbol{x}$. Note that $\nabla\boldsymbol{\mathcal{P}}\left(\boldsymbol{x}\right)$ is a third order tensor. Some more relations for the time derivatives of the projectors are given. To shorten the notation, the time-dependent projectors are rewritten as

$$\boldsymbol{\mathcal{P}}\left(t\right) = \boldsymbol{\mathcal{P}}\left(\boldsymbol{x}\left(t\right)\right), \qquad \boldsymbol{\mathcal{Q}}\left(t\right) = \boldsymbol{\mathcal{Q}}\left(\boldsymbol{x}\left(t\right)\right). \tag{A.77}$$

The time derivative is denoted as

$$\dot{\boldsymbol{\mathcal{P}}}\left(t\right) = \frac{d}{dt}\boldsymbol{\mathcal{P}}\left(\boldsymbol{x}\left(t\right)\right) = \nabla\boldsymbol{\mathcal{P}}\left(\boldsymbol{x}\left(t\right)\right)\dot{\boldsymbol{x}}\left(t\right). \tag{A.78}$$

From the complementarity property Eq. (2.6) follows

$$\dot{\boldsymbol{\mathcal{P}}}\left(t\right)\boldsymbol{\mathcal{Q}}\left(t\right) + \boldsymbol{\mathcal{P}}\left(t\right)\dot{\boldsymbol{\mathcal{Q}}}\left(t\right) = \boldsymbol{0}, \qquad \dot{\boldsymbol{\mathcal{Q}}}\left(t\right)\boldsymbol{\mathcal{P}}\left(t\right) + \boldsymbol{\mathcal{Q}}\left(t\right)\dot{\boldsymbol{\mathcal{P}}}\left(t\right) = \boldsymbol{0}, \tag{A.79}$$

$$\dot{\boldsymbol{\mathcal{Q}}}\left(t\right)\boldsymbol{\mathcal{Q}}\left(t\right) + \boldsymbol{\mathcal{Q}}\left(t\right)\dot{\boldsymbol{\mathcal{Q}}}\left(t\right) = \dot{\boldsymbol{\mathcal{Q}}}\left(t\right), \qquad \dot{\boldsymbol{\mathcal{P}}}\left(t\right)\boldsymbol{\mathcal{P}}\left(t\right) + \boldsymbol{\mathcal{P}}\left(t\right)\dot{\boldsymbol{\mathcal{P}}}\left(t\right) = \dot{\boldsymbol{\mathcal{P}}}\left(t\right). \tag{A.80}$$

The last line yields

$$\boldsymbol{\mathcal{Q}}\left(t\right)\dot{\boldsymbol{\mathcal{Q}}}\left(t\right)\boldsymbol{\mathcal{Q}}\left(t\right) + \boldsymbol{\mathcal{Q}}\left(t\right)\boldsymbol{\mathcal{Q}}\left(t\right)\dot{\boldsymbol{\mathcal{Q}}}\left(t\right) = \boldsymbol{\mathcal{Q}}\left(t\right)\dot{\boldsymbol{\mathcal{Q}}}\left(t\right), \tag{A.81}$$

or

$$\boldsymbol{\mathcal{Q}}\left(t\right)\dot{\boldsymbol{\mathcal{Q}}}\left(t\right)\boldsymbol{\mathcal{Q}}\left(t\right) = \boldsymbol{0}. \tag{A.82}$$

A similar computation results in

$$\boldsymbol{\mathcal{P}}\left(t\right)\dot{\boldsymbol{\mathcal{Q}}}\left(t\right)\boldsymbol{\mathcal{Q}}\left(t\right) + \boldsymbol{\mathcal{P}}\left(t\right)\boldsymbol{\mathcal{Q}}\left(t\right)\dot{\boldsymbol{\mathcal{Q}}}\left(t\right) = \boldsymbol{\mathcal{P}}\left(t\right)\dot{\boldsymbol{\mathcal{Q}}}\left(t\right), \tag{A.83}$$

or

$$\boldsymbol{\mathcal{P}}\left(t\right)\dot{\boldsymbol{\mathcal{Q}}}\left(t\right)\boldsymbol{\mathcal{Q}}\left(t\right) = \boldsymbol{\mathcal{P}}\left(t\right)\dot{\boldsymbol{\mathcal{Q}}}\left(t\right). \tag{A.84}$$

With the help of the projectors Eqs. (2.2), (2.3), the time derivative of $\boldsymbol{x}\left(t\right)$ can be written as

$$\begin{aligned}\frac{d}{dt}\boldsymbol{x}\left(t\right) &= \frac{d}{dt}\left(\boldsymbol{\mathcal{P}}\left(t\right)\boldsymbol{x}\left(t\right) + \boldsymbol{\mathcal{Q}}\left(t\right)\boldsymbol{x}\left(t\right)\right)\\ &= \dot{\boldsymbol{\mathcal{P}}}\left(t\right)\boldsymbol{x}\left(t\right) + \dot{\boldsymbol{\mathcal{Q}}}\left(t\right)\boldsymbol{x}\left(t\right) + \boldsymbol{\mathcal{P}}\left(t\right)\dot{\boldsymbol{x}}\left(t\right) + \boldsymbol{\mathcal{Q}}\left(t\right)\dot{\boldsymbol{x}}\left(t\right).\end{aligned} \tag{A.85}$$

Applying $\boldsymbol{\mathcal{Q}}\left(t\right)$ from the left and using Eq. (A.74) yields

$$\boldsymbol{\mathcal{Q}}\left(t\right)\dot{\boldsymbol{\mathcal{P}}}\left(t\right)\boldsymbol{x}\left(t\right) + \boldsymbol{\mathcal{Q}}\left(t\right)\dot{\boldsymbol{\mathcal{Q}}}\left(t\right)\boldsymbol{x}\left(t\right) + \boldsymbol{\mathcal{Q}}\left(t\right)\dot{\boldsymbol{x}}\left(t\right) = \boldsymbol{\mathcal{Q}}\left(t\right)\dot{\boldsymbol{x}}\left(t\right). \tag{A.86}$$

Similarly, applying $\boldsymbol{\mathcal{P}}\left(t\right)$ from the left gives

$$\boldsymbol{\mathcal{P}}\left(t\right)\dot{\boldsymbol{\mathcal{P}}}\left(t\right)\boldsymbol{x}\left(t\right) + \boldsymbol{\mathcal{P}}\left(t\right)\dot{\boldsymbol{\mathcal{Q}}}\left(t\right)\boldsymbol{x}\left(t\right) + \boldsymbol{\mathcal{P}}\left(t\right)\dot{\boldsymbol{x}}\left(t\right) = \boldsymbol{\mathcal{P}}\left(t\right)\dot{\boldsymbol{x}}\left(t\right). \tag{A.87}$$

A.4. Diagonalizing the projectors $\boldsymbol{\mathcal{P}}(\boldsymbol{x})$ and $\boldsymbol{\mathcal{Q}}(\boldsymbol{x})$

Let the $n \times n$ matrix $\boldsymbol{\mathcal{Q}}(\boldsymbol{x})$ be the projector

$$\boldsymbol{\mathcal{Q}}(\boldsymbol{x}) = \mathbf{1} - \boldsymbol{\mathcal{B}}(\boldsymbol{x})\,\boldsymbol{\mathcal{B}}^{+}(\boldsymbol{x}), \tag{A.88}$$

with $\boldsymbol{\mathcal{B}}^{+}(\boldsymbol{x})$ the Moore-Penrose pseudo inverse of the coupling matrix $\boldsymbol{\mathcal{B}}(\boldsymbol{x})$. $\boldsymbol{\mathcal{Q}}(\boldsymbol{x})$ is idempotent,

$$\boldsymbol{\mathcal{Q}}(\boldsymbol{x})\,\boldsymbol{\mathcal{Q}}(\boldsymbol{x}) = \boldsymbol{\mathcal{Q}}(\boldsymbol{x}), \tag{A.89}$$

and has a complementary projector defined by

$$\boldsymbol{\mathcal{P}}(\boldsymbol{x}) = \mathbf{1} - \boldsymbol{\mathcal{Q}}(\boldsymbol{x}). \tag{A.90}$$

Furthermore, $\boldsymbol{\mathcal{Q}}(\boldsymbol{x})$ satisfies

$$\boldsymbol{\mathcal{Q}}(\boldsymbol{x})\,\boldsymbol{\mathcal{B}}(\boldsymbol{x}) = \boldsymbol{\mathcal{B}}(\boldsymbol{x}) - \boldsymbol{\mathcal{B}}(\boldsymbol{x})\,\boldsymbol{\mathcal{B}}^{+}(\boldsymbol{x})\,\boldsymbol{\mathcal{B}}(\boldsymbol{x}) = \mathbf{0}. \tag{A.91}$$

Assume that the rank of $\boldsymbol{\mathcal{Q}}(\boldsymbol{x})$ is, with $p \leq n$,

$$\operatorname{rank}\left(\boldsymbol{\mathcal{Q}}(\boldsymbol{x})\right) = n - p, \tag{A.92}$$

for all $\boldsymbol{x}$.

Any projector $\boldsymbol{\mathcal{Q}}(\boldsymbol{x})$ can be diagonalized with zeros and ones as the diagonal entries (Fischer, 2013; Liesen and Mehrmann, 2015). Let $\boldsymbol{\mathcal{T}}(\boldsymbol{x})$ be the $n \times n$ matrix which diagonalizes the projector $\boldsymbol{\mathcal{Q}}(\boldsymbol{x})$,

$$\boldsymbol{\mathcal{Q}}_D = \boldsymbol{\mathcal{T}}^{-1}(\boldsymbol{x})\,\boldsymbol{\mathcal{Q}}(\boldsymbol{x})\,\boldsymbol{\mathcal{T}}(\boldsymbol{x}) = \begin{pmatrix} 0 & \cdots & 0 & \cdots & 0 \\ \vdots & \ddots & \vdots & \vdots & \vdots \\ 0 & \cdots & 1 & \cdots & \vdots \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & \cdots & 0 & \cdots & 1 \end{pmatrix}, \tag{A.93}$$

such that the first p diagonal elements are zero, while the last $n-p$ diagonal elements are one. The same matrix $\boldsymbol{\mathcal{T}}(\boldsymbol{x})$ diagonalizes the projector $\boldsymbol{\mathcal{P}}$ as well,

$$\begin{aligned} \boldsymbol{\mathcal{P}}_D &= \boldsymbol{\mathcal{T}}^{-1}(\boldsymbol{x})\,\boldsymbol{\mathcal{P}}\boldsymbol{\mathcal{T}}(\boldsymbol{x}) \\ &= \boldsymbol{\mathcal{T}}^{-1}(\boldsymbol{x})\,(\mathbf{1} - \boldsymbol{\mathcal{Q}})\,\boldsymbol{\mathcal{T}}(\boldsymbol{x}) = \mathbf{1} - \boldsymbol{\mathcal{Q}}_D = \begin{pmatrix} 1 & \cdots & 0 & \cdots & 0 \\ \vdots & \ddots & \vdots & \vdots & \vdots \\ 0 & \cdots & 1 & \cdots & \vdots \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & \cdots & 0 & \cdots & 0 \end{pmatrix}, \end{aligned} \tag{A.94}$$

such that the first p diagonal elements are one, while the last $n-p$ diagonal elements are zero. The matrix $\boldsymbol{\mathcal{T}}(\boldsymbol{x})$ contains n linearly independent eigenvectors of $\boldsymbol{\mathcal{Q}}(\boldsymbol{x})$ and can be constructed as follows. Let $\boldsymbol{q}_i(\boldsymbol{x})$ denote the i-th column of the matrix $\boldsymbol{\mathcal{Q}}(\boldsymbol{x})$,

$$\boldsymbol{\mathcal{Q}}(\boldsymbol{x}) = \Big(\ \boldsymbol{q}_1(\boldsymbol{x}), \quad \ldots, \quad \boldsymbol{q}_n(\boldsymbol{x}) \ \Big), \tag{A.95}$$

or, written component wise,

$$\mathcal{Q}_{ij}(\boldsymbol{x}) = \Big(\boldsymbol{q}_j\Big)_i(\boldsymbol{x}) = q_{j,i}(\boldsymbol{x}). \tag{A.96}$$

The vectors $\boldsymbol{q}_i(\boldsymbol{x})$ are eigenvectors of $\boldsymbol{\mathcal{Q}}(\boldsymbol{x})$. Indeed, $\boldsymbol{\mathcal{Q}}(\boldsymbol{x})$ is idempotent,

$$\boldsymbol{\mathcal{Q}}(\boldsymbol{x})\,\boldsymbol{\mathcal{Q}}(\boldsymbol{x}) = \boldsymbol{\mathcal{Q}}(\boldsymbol{x}), \tag{A.97}$$

or, written component wise,

$$\sum_{j=1}^{n} \mathcal{Q}_{ij}(\boldsymbol{x})\,\mathcal{Q}_{jk}(\boldsymbol{x}) = \mathcal{Q}_{ik}(\boldsymbol{x}). \tag{A.98}$$

Expressed in terms of the vectors $\boldsymbol{q}_i(\boldsymbol{x})$, the last relation becomes

$$\sum_{j=1}^{n} \mathcal{Q}_{ij}(\boldsymbol{x})\,q_{k,j}(\boldsymbol{x}) = q_{k,i}(\boldsymbol{x}), \tag{A.99}$$

or

$$\boldsymbol{\mathcal{Q}}(\boldsymbol{x})\,\boldsymbol{q}_k(\boldsymbol{x}) = \boldsymbol{q}_k(\boldsymbol{x}), \tag{A.100}$$

which shows that the vectors $\boldsymbol{q}_k(\boldsymbol{x})$ are eigenvectors of $\boldsymbol{\mathcal{Q}}(\boldsymbol{x})$ to eigenvalue one. However, because $\boldsymbol{\mathcal{Q}}(\boldsymbol{x})$ has $\operatorname{rank}(\boldsymbol{\mathcal{Q}}(\boldsymbol{x})) = n - p$, only $n - p$ vectors out of $i = 1, \ldots, n$ vectors $\boldsymbol{q}_i(\boldsymbol{x})$ are linearly independent. By appropriately ordering the eigenvectors, one can ensure that the first $n - p$ eigenvectors $\boldsymbol{q}_1(\boldsymbol{x}), \ldots, \boldsymbol{q}_{n-p}(\boldsymbol{x})$ are linearly independent. The remaining eigenvectors can be constructed from the coupling matrix $\boldsymbol{\mathcal{B}}(\boldsymbol{x})$. The $n \times p$ matrix $\boldsymbol{\mathcal{B}}(\boldsymbol{x})$ can be written in terms of its p column vectors as

$$\boldsymbol{\mathcal{B}}(\boldsymbol{x}) = \Big(\ \boldsymbol{b}_1(\boldsymbol{x}), \quad \ldots, \quad \boldsymbol{b}_p(\boldsymbol{x}) \ \Big), \tag{A.101}$$

or, written component wise,

$$\mathcal{B}_{ij}(\boldsymbol{x}) = (\boldsymbol{b}_j)_i(\boldsymbol{x}) = b_{j,i}(\boldsymbol{x}). \tag{A.102}$$

From the relation

$$\boldsymbol{\mathcal{Q}}(\boldsymbol{x})\,\boldsymbol{\mathcal{B}}(\boldsymbol{x}) = \boldsymbol{0}, \tag{A.103}$$

or

$$\sum_{j=1}^{n} \mathcal{Q}_{ij}(\boldsymbol{x})\, \mathcal{B}_{jk}(\boldsymbol{x}) = 0, \tag{A.104}$$

follows that the vectors $\boldsymbol{b}_i(\boldsymbol{x})$ are indeed eigenvectors to eigenvalue zero,

$$\sum_{j=1}^{n} \mathcal{Q}_{ij}(\boldsymbol{x})\, b_{k,j}(\boldsymbol{x}) = 0, \tag{A.105}$$

or

$$\boldsymbol{\mathcal{Q}}(\boldsymbol{x})\, \boldsymbol{b}_k(\boldsymbol{x}) = 0. \tag{A.106}$$

Finally, the matrix $\boldsymbol{\mathcal{T}}(\boldsymbol{x})$ becomes

$$\boldsymbol{\mathcal{T}}(\boldsymbol{x}) = \left(\boldsymbol{b}_1(\boldsymbol{x}),\ \ldots,\ \boldsymbol{b}_p(\boldsymbol{x}),\ \boldsymbol{q}_1(\boldsymbol{x}),\ \ldots,\ \boldsymbol{q}_{n-p}(\boldsymbol{x}) \right). \tag{A.107}$$

If the projectors $\boldsymbol{\mathcal{P}}(\boldsymbol{x})$ and $\boldsymbol{\mathcal{Q}}(\boldsymbol{x})$ are not diagonal, splitting up the vector $\boldsymbol{x}(t)$ as

$$\boldsymbol{x}(t) = \boldsymbol{\mathcal{P}}(\boldsymbol{x}(t))\, \boldsymbol{x}(t) + \boldsymbol{\mathcal{Q}}(\boldsymbol{x}(t))\, \boldsymbol{x}(t) \tag{A.108}$$

results in the parts $\boldsymbol{\mathcal{P}}(\boldsymbol{x}(t))\, \boldsymbol{x}(t)$ and $\boldsymbol{\mathcal{Q}}(\boldsymbol{x}(t))\, \boldsymbol{x}(t)$ being nonlinear combinations of the components of $\boldsymbol{x}(t)$. Due to this nonlinear mixing, it is not clear which state components belong to which part. If the projectors are diagonal,

$$\boldsymbol{x}(t) = \boldsymbol{\mathcal{P}}_D \boldsymbol{x}(t) + \boldsymbol{\mathcal{Q}}_D \boldsymbol{x}(t), \tag{A.109}$$

the parts $\boldsymbol{\mathcal{P}}_D \boldsymbol{x}(t)$ and $\boldsymbol{\mathcal{Q}}_D \boldsymbol{x}(t)$ are linear combinations of the state components. Furthermore, only the first p components of $\boldsymbol{\mathcal{P}}_D \boldsymbol{x}(t)$ and the last $n-p$ components of $\boldsymbol{\mathcal{Q}}_D \boldsymbol{x}(t)$ are nonzero, and all other components vanish. Thus, diagonal projectors allow a clear interpretation which components of $\boldsymbol{x}(t)$ belong to which part. If the projectors $\boldsymbol{\mathcal{P}}(\boldsymbol{x})$ and $\boldsymbol{\mathcal{Q}}(\boldsymbol{x})$ are not diagonal, the matrix $\boldsymbol{\mathcal{T}}(\boldsymbol{x})$ defines a coordinate transformation as follows. Let the vector $\boldsymbol{y}(t)$ be defined by

$$\boldsymbol{y}(t) = \boldsymbol{\mathcal{T}}^{-1}(\boldsymbol{x}(t))\, \boldsymbol{x}(t). \tag{A.110}$$

According to the construction above, the matrix $\boldsymbol{\mathcal{T}}(\boldsymbol{x}(t))$ always exists and is invertible. The inverse relation of Eq. (A.110) is

$$\boldsymbol{x}(t) = \boldsymbol{\mathcal{T}}(\boldsymbol{x}(t))\, \boldsymbol{y}(t). \tag{A.111}$$

Splitting up $\boldsymbol{x}(t)$ in Eq. (A.110) with the help of the projectors $\boldsymbol{\mathcal{P}}(\boldsymbol{x}(t))$ and $\boldsymbol{\mathcal{Q}}(\boldsymbol{x}(t))$ gives

$$\begin{aligned}
\boldsymbol{y}(t) &= \boldsymbol{\mathcal{T}}^{-1}(\boldsymbol{x}(t))\, \boldsymbol{x}(t) \\
&= \boldsymbol{\mathcal{T}}^{-1}(\boldsymbol{x}(t))\, \boldsymbol{\mathcal{P}}(\boldsymbol{x}(t))\, \boldsymbol{x}(t) + \boldsymbol{\mathcal{T}}^{-1}(\boldsymbol{x}(t))\, \boldsymbol{\mathcal{Q}}(\boldsymbol{x}(t))\, \boldsymbol{x}(t) \\
&= \boldsymbol{\mathcal{T}}^{-1}(\boldsymbol{x}(t))\, \boldsymbol{\mathcal{P}}(\boldsymbol{x}(t))\, \boldsymbol{\mathcal{T}}(\boldsymbol{x}(t))\, \boldsymbol{y}(t) + \boldsymbol{\mathcal{T}}^{-1}(\boldsymbol{x}(t))\, \boldsymbol{\mathcal{Q}}(\boldsymbol{x}(t))\, \boldsymbol{\mathcal{T}}(\boldsymbol{x}(t))\, \boldsymbol{y}(t) \\
&= \boldsymbol{\mathcal{P}}_D \boldsymbol{y}(t) + \boldsymbol{\mathcal{Q}}_D \boldsymbol{y}(t).
\end{aligned} \tag{A.112}$$

Thus, in the new coordinates, the state can be separated in two parts $\boldsymbol{\mathcal{P}}_D\boldsymbol{y}\,(t)$ and $\boldsymbol{\mathcal{Q}}_D\boldsymbol{y}\,(t)$ which are linear combinations of the state components of $\boldsymbol{y}$. The first p components of $\boldsymbol{y}$ belong to $\boldsymbol{\mathcal{P}}_D\boldsymbol{y}\,(t)$ and the last $n-p$ components of $\boldsymbol{y}$ belong to $\boldsymbol{\mathcal{Q}}_D\boldsymbol{y}\,(t)$. Such a representation can be viewed as a normal form suitable for computations with affine control systems. Note that Eq. (A.110) yields an explicit expression for the new coordinates $\boldsymbol{y}$ in terms of the old coordinates $\boldsymbol{x}$. To obtain $\boldsymbol{x}$ in terms of $\boldsymbol{y}$, Eq. (A.110) must be solved for $\boldsymbol{x}$.

Bibliography

A. Albert. *Regression and the Moore-Penrose Pseudoinverse*. Academic Press, 1972. ISBN 9780124109582.

S. Alonso, J. Löber, M. Bär, and H. Engel. Wave propagation in heterogeneous bistable and excitable media. *Eur. Phys. J. ST*, 187(1):31–40, 2010. doi:10.1140/epjst/e2010-01268-1.

D. Anderson, G. B. McFadden, and A. Wheeler. Diffuse-interface methods in fluid mechanics. *Annu. Rev. Fluid Mech.*, 30(1):139–165, 1998. doi:10.1146/annurev.fluid.30.1.139.

J. Angeles. *Fundamentals of Robotic Mechanical Systems: Theory, Methods, and Algorithms*. Number 124 in Mechanical Engineering Series. Springer, 4th edition, 2013. ISBN 9783319018508.

I. S. Aranson, J. Löber, and F. Ziebert. Phase-field description of substrate-based motility of eukaryotic cells. In A. S. Mikhailov and G. Ertl, editors, *Engineering of Chemical Complexity II*, World Scientific Lecture Notes in Complex Systems, pages 93–104. World Scientific, 2014. ISBN 9814390453. doi:10.1142/9789814616133_0006.

N. T. Bailey. *The Mathematical Theory of Infectious Diseases*. Hafner Press/ MacMillian Pub. Co., 2nd edition, 1975. ISBN 9780852642313.

R. C. Balescu. *Equilibrium and Non-Equilibrium Statistical Mechanics*. Wiley, 1st edition, 1975. ISBN 9780471046004.

D. J. Bell and D. H. Jacobson. *Singular Optimal Control Problems*. Number 117 in Mathematics in Science and Engineering. Elsevier Science, 1975. ISBN 9780120850600.

R. Bellman. *Dynamic Programming*. Dover Publications, Reprint edition, 2003. ISBN 9780486428093.

C. M. Bender and S. A. Orszag. *Advanced Mathematical Methods for Scientists and Engineers I: Asymptotic Methods and Perturbation Theory*. Springer, 2010. ISBN 9781441931870.

W. Boettinger, J. Warren, C. Beckermann, and A. Karma. Phase-field simulation of solidification. *Annu. Rev. Mat. Res.*, 32(1):163–194, 2002. doi:10.1146/annurev.matsci.32.101901.155803.

J. A. E. Bryson and Y.-C. Ho. *Applied Optimal Control: Optimization, Estimation and Control*. CRC Press, Revised edition, 1975. ISBN 9780891162285.

R. Buchholz, H. Engel, E. Kammann, and F. Tröltzsch. On the optimal control of the Schlögl-model. *Computational Optimization and Applications*, 56(1):153–185, 2013. doi:10.1007/s10589-013-9550-y.

S. L. Campbell. *Singular Systems of Differential Equations I*. Chapman & Hall/CRC Research Notes in Mathematics Series. Pitman Publishing, 1980. ISBN 9780273084389.

S. L. Campbell. *Singular Systems of Differential Equations II*. Chapman & Hall/CRC Research Notes in Mathematics Series. Pitman Publishing, 1982. ISBN 9780273085164.

S. L. Campbell and C. D. Meyer Jr. *Generalized Inverses of Linear Transformations*. Dover Publications, 1991. ISBN 9780486666938.

E. Casas and F. Tröltzsch. Second-order and stability analysis for state-constrained elliptic optimal control problems with sparse controls. *SIAM Journal on Control and Optimization*, 52(2):1010–1033, 2014. doi:10.1137/130917314.

C.-T. Chen. *Linear System Theory and Design*. Oxford Series in Electrical and Computer Engineering. Oxford University Press, 3rd edition, 1998. ISBN 9780195117776.

C. Cohen-Tannoudji, B. Diu, and F. Laloe. *Quantenmechanik*. Walter De Gruyter, 4th edition, 2010. ISBN 9783110241136.

M. Cross and P. Hohenberg. Pattern formation outside of equilibrium. *Rev. Mod. Phys.*, 65(3):851–1112, 1993. doi:10.1103/RevModPhys.65.851.

C. C. de Wit, B. Siciliano, and G. Bastin, editors. *Theory of Robot Control*. Communications and Control Engineering. Springer, 1st edition, 2012. ISBN 9781447115038.

A. Engel. Noise-induced front propagation in a bistable system. *Phys. Lett. A*, 113 (3):139–142, 1985. doi:10.1016/0375-9601(85)90157-4.

A. Engel and W. Ebeling. Interaction of moving interfaces with obstacles. *Phys. Lett. A*, 122(1):20–24, 1987. doi:10.1016/0375-9601(87)90768-7.

I. R. Epstein and J. A. Pojman. *An Introduction to Nonlinear Chemical Dynamics: Oscillations, Waves, Patterns, and Chaos*. Topics in Physical Chemistry. Oxford University Press, 1st edition, 1998. ISBN 9780195096705.

E. Fick. *Einführung in die Grundlagen der Quantenmechanik*. Aula Verlag, 6th edition, 1988. ISBN 3891044720.

R. Field and R. Noyes. Oscillations in chemical systems. IV. Limit cycle behavior in a model of a real chemical reaction. *J. Chem. Phys.*, 60:1877, 1974. doi:10.1063/1.1681288.

R. Field, E. Körös, and R. Noyes. Oscillations in chemical systems. II. Thorough analysis of temporal oscillation in the bromate-cerium-malonic acid system. *J. Am. Chem. Soc.*, 94(25):8649–8664, 1972. doi:10.1021/ja00780a001.

G. Fischer. *Lineare Algebra: Eine Einführung für Studienanfänger*. Springer Spektrum, 18th edition, 2013. ISBN 9783658039448.

R. FitzHugh. Impulses and physiological states in theoretical models of nerve membrane. *Biophysical J.*, 1(6):445–466, 1961. doi:10.1016/S0006-3495(61)86902-6.

M. Fliess, J. Lévine, P. Martin, and P. Rouchon. Flatness and defect of non-linear systems: Introductory theory and examples. *Int. J. Control*, 61(6):1327–1361, 1995. doi:10.1080/00207179508921959.

R. A. Freeman and P. V. Kokotovic. *Robust Nonlinear Control Design: State-Space and Lyapunov Techniques*. Systems & Control: Foundations & Applications. Birkhäuser Boston, 1st edition, 1996. ISBN 0817647589.

H. Goldstein, C. P. Poole Jr., and J. L. Safko. *Classical Mechanics*. Addison-Wesley, 3rd edition, 2001. ISBN 9780201657029.

H. Grabert. *Projection Operator Techniques in Nonequilibrium Statistical Mechanics*. Number 95 in Springer Tracts in Modern Physics. Springer-Verlag, 1st edition, 1982. ISBN 9780387116358.

A. Hagberg and E. Meron. From labyrinthine patterns to spiral turbulence. *Phys. Rev. Lett.*, 72:2494–2497, 1994. doi:10.1103/PhysRevLett.72.2494.

A. L. Hodgkin and A. F. Huxley. A quantitative description of membrane current and its application to conduction and excitation in nerve. *J. Physiol.*, 117(4):500, 1952. doi:10.1113/jphysiol.1952.sp004764.

B. Houska, H. Ferreau, and M. Diehl. ACADO Toolkit – An Open Source Framework for Automatic Control and Dynamic Optimization. *Optimal Control Applications and Methods*, 32(3):298–312, 2011a. doi:10.1002/oca.939.

B. Houska, H. Ferreau, and M. Diehl. An Auto-Generated Real-Time Iteration Algorithm for Nonlinear MPC in the Microsecond Range. *Automatica*, 47(10): 2279–2285, 2011b. doi:10.1016/j.automatica.2011.08.020.

B. Houska, H. Ferreau, M. Vukov, and R. Quirynen. *ACADO Toolkit User's Manual*, 2013. URL `http://www.acadotoolkit.org`.

D. G. Hull. *Optimal Control Theory for Applications*. Mechanical Engineering Series. Springer, 1st edition, 2003. ISBN 9780387400709.

A. Isidori. *Nonlinear Control Systems*. Communications and Control Engineering. Springer, 3rd edition, 1995. ISBN 9783540199168.

E. M. Izhikevich. *Dynamical Systems in Neuroscience: The Geometry of Excitability and Bursting*. Computational Neuroscience. The MIT Press, 2010. ISBN 9780262514200.

R. Johnson. *Singular Perturbation Theory: Mathematical and Analytical Techniques with Applications to Engineering*. Springer, 1st edition, 2004. ISBN 9780387232003.

W. Just, A. Pelster, M. Schanz, and E. Schöll. Delayed complex systems: an overview. *Philos. T. Roy. Soc. A*, 368(1911):303–304, 2010. doi:10.1098/rsta.2009.0243.

T. Kailath. *Linear Systems*. Prentice-Hall, Inc., 1st edition, 1980. ISBN 9780135369616.

R. Kalman. On the general theory of control systems. *IEEE Trans. Autom. Control*, 4(3):110–110, 1959. doi:10.1109/TAC.1959.1104873.

R. E. Kalman. Contributions to the theory of optimal control. *Bol. Soc. Mat. Mexicana*, 5(2):102–119, 1960.

H. J. Kappen. Linear theory for control of nonlinear stochastic systems. *Phys. Rev. Lett.*, 95:200201, 2005. doi:10.1103/PhysRevLett.95.200201.

R. Kapral and K. Showalter, editors. *Chemical Waves and Patterns*. Springer, 1995. ISBN 9789401045049.

J. Keener and J. Sneyd. *Mathematical Physiology: I: Cellular Physiology*. Number 8 in Interdisciplinary Applied Mathematics. Springer, 2nd edition, 2008a. ISBN 9780387758466.

J. Keener and J. Sneyd. *Mathematical Physiology: II: Systems Physiology*. Number 8 in Interdisciplinary Applied Mathematics. Springer, 2nd edition, 2008b. ISBN 9780387793870.

W. O. Kermack and A. G. McKendrick. A contribution to the mathematical theory of epidemics. *P. Roy. Soc. Lond. A: Mat.*, 115(772):700–721, 1927. doi:10.1098/rspa.1927.0118.

P. Kevrekidis, I. Kevrekidis, B. Malomed, H. Nistazakis, and D. Frantzeskakis. Dragging bistable fronts. *Phys. Scr.*, 69(6):451, 2004. doi:10.1238/Physica.Regular.069a00451.

H. K. Khalil. *Nonlinear Systems*. Prentice Hall, 3rd edition, 2001. ISBN 9780130673893.

H. J. Krug, L. Pohlmann, and L. Kuhnert. Analysis of the modified complete oregonator accounting for oxygen sensitivity and photosensitivity of Belousov-Zhabotinskii systems. *J. Phys. Chem.*, 94(12):4862–4866, 1990. doi:10.1021/j100375a021.

P. Kunkel and V. Mehrmann. *Differential-Algebraic Equations: Analysis and Numerical Solution*. European Mathematical Society, 2006. ISBN 9783037190173.

Y. Kuramoto. *Chemical Oscillations, Waves, and Turbulence*. Dover Books on Chemistry. Dover Publications, 2003. ISBN 9780486428819.

D. Lebiedz and U. Brandt-Pollmann. Manipulation of self-aggregation patterns and waves in a reaction-diffusion system by optimal boundary control strategies. *Phys. Rev. Lett.*, 91:208301, 2003. doi:10.1103/PhysRevLett.91.208301.

J. Levine. *Analysis and Control of Nonlinear Systems: A Flatness-based Approach*. Mathematical Engineering. Springer, 1st edition, 2009. ISBN 9783642008382.

F. L. Lewis, C. T. Abdallah, and D. M. Dawson. *Control of Robot Manipulators*. Macmillan Coll Div, 1st edition, 1993. ISBN 9780023705014.

D. Liberzon. *Calculus of Variations and Optimal Control Theory: A Concise Introduction*. Princeton University Press, 2012. ISBN 9780691151878.

J. Liesen and V. Mehrmann. *Lineare Algebra: Ein Lehrbuch über die Theorie mit Blick auf die Praxis*. Springer Spektrum, 2nd edition, 2015. ISBN 9783658066093.

J. Löber. Nonlinear excitation waves in spatially heterogeneous reaction-diffusion systems. Diploma thesis, Technical University of Berlin, 2009.

J. Löber. Stability of position control of traveling waves in reaction-diffusion systems. *Phys. Rev. E*, 89:062904, 2014. doi:10.1103/PhysRevE.89.062904.

J. Löber and H. Engel. Analytical approximations for spiral waves. *Chaos*, 23(4): 043135, 2013. doi:10.1063/1.4848576.

J. Löber and H. Engel. Controlling the position of traveling waves in reaction-diffusion systems. *Phys. Rev. Lett.*, 112:148305, 2014. doi:10.1103/PhysRevLett.112.148305.

J. Löber, M. Bär, and H. Engel. Front propagation in one-dimensional spatially periodic bistable media. *Phys. Rev. E*, 86:066210, 2012. doi:10.1103/PhysRevE.86.066210.

J. Löber, R. Coles, J. Siebert, H. Engel, and E. Schöll. Control of chemical wave propagation. In A. S. Mikhailov and G. Ertl, editors, *Engineering of Chemical Complexity II*, World Scientific Lecture Notes in Complex Systems, pages 185–207. World Scientific, 2014. ISBN 9814390453. doi:10.1142/9789814616133_0011.

J. Löber, S. Martens, and H. Engel. Shaping wave patterns in reaction-diffusion systems. *Phys. Rev. E*, 90:062911, 2014. doi:10.1103/PhysRevE.90.062911.

J. Löber, F. Ziebert, and I. S. Aranson. Modeling crawling cell movement on soft engineered substrates. *Soft Matter*, 10:1365–1373, 2014. doi:10.1039/C3SM51597D.

J. Löber, F. Ziebert, and I. S. Aranson. Collisions of deformable cells lead to collective migration. *Sci. Rep.*, 5:9172, 2015. doi:10.1038/srep09172.

S. Luther, F. H. Fenton, B. G. Kornreich, A. Squires, P. Bittihn, D. Hornung, M. Zabel, J. Flanders, A. Gladuli, L. Campoy, et al. Low-energy control of electrical turbulence in the heart. *Nature*, 475(7355):235–239, 2011. doi:10.1038/nature10216.

B. A. Malomed, D. J. Frantzeskakis, H. E. Nistazakis, A. N. Yannacopoulos, and P. G. Kevrekidis. Pulled fronts in the Cahn–Hilliard equation. *Phys. Lett. A*, 295 (5):267–272, 2002. doi:10.1016/S0375-9601(02)00173-1.

S. Martens, J. Löber, and H. Engel. Front propagation in channels with spatially modulated cross section. *Phys. Rev. E*, 91:022902, 2015. doi:10.1103/PhysRevE.91.022902.

E. Mihaliuk, T. Sakurai, F. Chirila, and K. Showalter. Feedback stabilization of unstable propagating waves. *Phys. Rev. E*, 65(6):065602–65602, 2002. doi:10.1103/PhysRevE.65.065602.

A. Mikhailov. *Foundations of Synergetics I: Distributed Active Systems*. Number 51 in Springer Series in Synergetics. Springer, 2nd edition, 2011. ISBN 9783642785580.

A. S. Mikhailov and K. Showalter. Control of waves, patterns and turbulence in chemical systems. *Phys. Rep.*, 425(2):79–194, 2006. doi:10.1016/j.physrep.2005.11.003.

S. Molnos, J. Löber, J. F. Totz, and H. Engel. Control of transversal instabilities in reaction-diffusion systems. *preprint*, 2015. arXiv:1501.03769.

J. D. Murray. *Mathematical Biology: I. An Introduction*. Number 17 in Interdisciplinary Applied Mathematics. Springer, 3rd edition, 2007. ISBN 9780387952239.

J. D. Murray. *Mathematical Biology II: Spatial Models and Biomedical Applications*. Number 18 in Interdisciplinary Applied Mathematics. Springer, 3rd edition, 2011. ISBN 9780387952284.

J. Nagumo, S. Arimoto, and S. Yoshizawa. An active pulse transmission line simulating nerve axon. *Proc. IRE*, 50(10):2061–2070, 1962. doi:10.1109/JRPROC.1962.288235.

H. E. Nistazakis, P. G. Kevrekidis, B. A. Malomed, D. J. Frantzeskakis, and A. R. Bishop. Targeted transfer of solitons in continua and lattices. *Phys. Rev. E*, 66: 015601, 2002. doi:10.1103/PhysRevE.66.015601.

J. Nocedal and S. Wright. *Numerical Optimization*. Springer Series in Operations Research and Financial Engineering. Springer, 2nd edition, 2006. ISBN 9780387303031.

K. Ogata. *Modern Control Engineering*. Prentice Hall, 5th edition, 2009. ISBN 9780136156734.

E. Ott, C. Grebogi, and J. A. Yorke. Controlling chaos. *Phys. Rev. Lett.*, 64: 1196–1199, 1990. doi:10.1103/PhysRevLett.64.1196.

P. V. Paulau, J. Löber, and H. Engel. Stabilization of a scroll ring by a cylindrical neumann boundary. *Phys. Rev. E*, 88:062917, 2013. doi:10.1103/PhysRevE.88.062917.

L. Pismen. *Patterns and Interfaces in Dissipative Dynamics*. Springer Series in Synergetics. Springer, 2006. ISBN 9783540304302.

L. S. Pontryagin and V. G. Boltyanskii. *The Mathematical Theory of Optimal Processes*. John Wiley & Sons Inc, 1st edition, 1962. ISBN 9780470693810.

K. Pyragas. Continuous control of chaos by self-controlling feedback. *Phys. Lett. A*, 170(6):421 – 428, 1992. doi:10.1016/0375-9601(92)90745-8.

K. Pyragas. Delayed feedback control of chaos. *Philos. T. Roy. Soc. A*, 364(1846): 2309–2334, 2006. doi:10.1098/rsta.2006.1827.

C. Ryll. Numerische Analysis für Sparse Controls bei semilinearen parabolischen Differentialgleichungen. Master's thesis, Technical University of Berlin, 2011.

T. Sakurai, E. Mihaliuk, F. Chirila, and K. Showalter. Design and control of wave propagation patterns in excitable media. *Science*, 296(5575):2009–2012, 2002. doi:10.1126/science.1071265.

B. Sandstede. Stability of travelling waves. In B. Fiedler, editor, *Handbook of Dynamical Systems, Volume 2*, pages 983 – 1055. Elsevier Science, 2002. ISBN 0444501681. doi:10.1016/S1874-575X(02)80039-X.

L. Schimansky-Geier, A. Mikhailov, and W. Ebeling. Effect of fluctuation on plane front propagation in bistable nonequilibrium systems. *Ann. Phys. (Leipzig)*, 40 (4-5):277–286, 1983. doi:10.1002/andp.19834950412.

L. Schimansky-Geier, B. Fiedler, J. Kurths, and E. Schöll, editors. *Analysis and Control of Complex Nonlinear Processes in Physics, Chemistry and Biology*. Number 5 in World Scientific Lecture Notes in Complex Systems. World Scientific Pub Co Inc, 2007. ISBN 9789812705839.

J. Schlesner, V. Zykov, H. Engel, and E. Schöll. Stabilization of unstable rigid rotation of spiral waves in excitable media. *Phys. Rev. E*, 74(4):046215, 2006. doi:10.1103/PhysRevE.74.046215.

J. Schlesner, V. Zykov, H. Brandtstädter, I. Gerdes, and H. Engel. Efficient control of spiral wave location in an excitable medium with localized heterogeneities. *New J. Phys.*, 10(1):015003, 2008. doi:10.1088/1367-2630/10/1/015003.

F. Schlögl. Chemical reaction models for non-equilibrium phase transitions. *Z. Phys. A*, 253(2):147–161, 1972. doi:10.1007/BF01379769.

E. Schöll and H. G. Schuster, editors. *Handbook of Chaos Control*. Wiley-VCH, 2nd edition, 2007. ISBN 9783527406050.

J. R. Shewchuk. An introduction to the conjugate gradient method without the agonizing pain. Technical report, Carnegie Mellon University, Pittsburgh, PA, USA, 1994. URL `http://www.cs.cmu.edu/~quake-papers/painless-conjugate-gradient.pdf`.

T. Shinbrot, C. Grebogi, E. Ott, and J. A. Yorke. Using small perturbations to control chaos. *Nature*, 363(6428):411–417, 1993. doi:10.1038/363411a0.

H. Sira-Ramírez and S. K. Agrawal. *Differentially Flat Systems*. Number 17 in Automation and Control Engineering. CRC Press, 1st edition, 2004. ISBN 9780824754709.

J.-J. Slotine and W. Li. *Applied Nonlinear Control*. Prentice Hall, 1991. ISBN 9780130408907.

E. D. Sontag. Stability and feedback stabilization. In R. A. Meyers, editor, *Mathematics of Complexity and Dynamical Systems*, pages 1639–1652. Springer New York, 2011. ISBN 9781461418054. doi:10.1007/978-1-4614-1806-1_105.

O. Steinbock, V. Zykov, and S. Müller. Control of spiral-wave dynamics in active media by periodic modulation of excitability. *Nature*, 366(6453):322–324, 1993. doi:10.1038/366322a0.

K. Theißen. *Optimale Steuerprozesse unter partiellen Differentialgleichungs-Restriktionen mit linear eingehender Steuerfunktion*. PhD thesis, Westfälische Wilhelms-Universität Münster, Münster, Germany, 2006.

F. Tröltzsch. *Optimal Control of Partial Differential Equations*. Number 112 in Graduate Studies in Mathematics. American Mathematical Society, 2010. ISBN 9780821849040.

A. M. Turing. The chemical basis of morphogenesis. *Philos. T. Roy. Soc. B*, 237 (641):37–72, 1952. doi:10.1098/rstb.1952.0012.

J. J. Tyson and J. P. Keener. Singular perturbation theory of traveling waves in excitable media (a review). *Physica D*, 32(3):327–361, 1988. doi:10.1016/0167-2789(88)90062-0.

B. Van der Pol. On "relaxation-oscillations". *Lond. Edinb. Dubl. Phil. Mag.*, 2(11): 978–992, 1926. doi:10.1080/14786442608564127.

M. J. Van Nieuwstadt and R. M. Murray. Real time trajectory generation for differentially flat systems. Technical report, California Institute of Technology, 1997. URL `http://resolver.caltech.edu/CaltechCDSTR:1997.CIT-CDS-96-017`.

V. Vanag and I. Epstein. Design and control of patterns in reaction-diffusion systems. *Chaos*, 18(2):026107–026107, 2008. doi:10.1063/1.2900555.

V. K. Vanag and I. R. Epstein. Localized patterns in reaction-diffusion systems. *Chaos*, 17(3):037110, 2007. doi:10.1063/1.2752494.

R. Vinter. *Optimal Control*. Systems & Control: Foundations & Applications. Birkhäuser Boston, 1st edition, 2000. ISBN 9780817640750.

J. von Hardenberg, E. Meron, M. Shachak, and Y. Zarmi. Diversity of vegetation patterns and desertification. *Phys. Rev. Lett.*, 87:198101, 2001. doi:10.1103/PhysRevLett.87.198101.

J. Wolff, A. G. Papathanasiou, I. G. Kevrekidis, H. H. Rotermund, and G. Ertl. Spatiotemporal addressing of surface activity. *Science*, 294(5540):134–137, 2001. doi:10.1126/science.1063597.

J. Wolff, A. Papathanasiou, H. Rotermund, G. Ertl, M. Katsoulakis, X. Li, and I. Kevrekidis. Wave initiation through spatiotemporally controllable perturbations. *Phys. Rev. Lett.*, 90(14):148301, 2003a. doi:10.1103/PhysRevLett.90.148301.

J. Wolff, A. G. Papathanasiou, H. H. Rotermund, G. Ertl, X. Li, and I. G. Kevrekidis. Gentle dragging of reaction waves. *Phys. Rev. Lett.*, 90(1):018302, 2003b. doi:10.1103/PhysRevLett.90.018302.

Wolfram Research, Inc. Mathematica 10.0, 2014. URL `http://www.wolfram.com/mathematica`.

V. F. Zaitsev and A. D. Polyanin. *Handbook of Exact Solutions for Ordinary Differential Equations*. Chapman and Hall/CRC, 2nd edition, 2002. ISBN 9781584882978.

Y. B. Zeldovich and D. A. Frank-Kamenetskii. On the theory of uniform flame propagation. *Dokl. Akad. Nauk SSSR*, 19:693–798, 1938.

F. Ziebert and I. S. Aranson. Effects of adhesion dynamics and substrate compliance on the shape and motility of crawling cells. *PloS ONE*, 8(5):e64511, 2013. doi:10.1371/journal.pone.0064511.

F. Ziebert, S. Swaminathan, and I. S. Aranson. Model for self-polarization and motility of keratocyte fragments. *J. R. Soc. Interface*, page 20110433, 2011. doi:10.1098/rsif.2011.0433.

V. Zykov and H. Engel. Feedback-mediated control of spiral waves. *Physica D*, 199 (1):243–263, 2004. doi:10.1016/j.physd.2004.10.001.

V. Zykov, O. Steinbock, and S. Müller. External forcing of spiral waves. *Chaos*, 4 (3):509–518, 1994. doi:10.1063/1.166029.

V. S. Zykov. *Simulation of Wave Processes in Excitable Media*. Palgrave Macmillan, 1988. ISBN 9780719024726.

V. S. Zykov, G. Bordiougov, H. Brandtstädter, I. Gerdes, and H. Engel. Global control of spiral wave dynamics in an excitable domain of circular and elliptical shape. *Phys. Rev. Lett.*, 92:018304, 2004. doi:10.1103/PhysRevLett.92.018304.

Symbols

n	dimensionality of state space
p	dimension of vector of control signals
m	dimension of vector of output
t	time
t_0	initial time
t_1	terminal time
$\boldsymbol{x}(t)$	n-dimensional time-dependent state vector
$\boldsymbol{x}_d(t)$	n-dimensional desired trajectory
$\Delta\boldsymbol{x}(t)$	difference between state and desired trajectory
$\boldsymbol{x}_0$	n-dimensional vector of initial states
$\boldsymbol{x}_1$	n-dimensional vector of terminal states
$\boldsymbol{u}(t)$	p-dimensional vector of control or input signals
$\boldsymbol{\mathcal{A}}$	$n \times n$ state matrix
$\boldsymbol{R}(\boldsymbol{x})$	nonlinear kinetics of an n-dimensional dynamical system
$\boldsymbol{\mathcal{B}}(\boldsymbol{x})$	$n \times p$ input or coupling matrix
$\boldsymbol{\mathcal{P}}$	$n \times n$ projector
$\boldsymbol{\mathcal{Q}}$	$n \times n$ projector
$\boldsymbol{\mathcal{A}}^T$	transpose of matrix $\boldsymbol{\mathcal{A}}$
$\boldsymbol{\mathcal{A}}^+$	Moore-Penrose pseudo inverse of matrix $\boldsymbol{\mathcal{A}}$
$\boldsymbol{\mathcal{A}}^g$	generalized inverse of matrix $\boldsymbol{\mathcal{A}}$
$\nabla\boldsymbol{R}(\boldsymbol{x})$	$n \times n$ Jacobi matrix of $\boldsymbol{R}(\boldsymbol{x})$

$\boldsymbol{\Phi}(t, t_0)$	$n \times n$ state transition matrix for a linear dynamical system
$\boldsymbol{\mathcal{K}}$	Kalman's $n \times np$ controllability matrix
$\tilde{\boldsymbol{\mathcal{K}}}$	$n \times n^2$ controllability matrix for exactly realizable trajectories
$\tilde{\boldsymbol{\mathcal{K}}}_{\boldsymbol{\mathcal{N}}}$	$n \times (n^2 + n)$ output trajectory realizability matrix
$\boldsymbol{\mathcal{C}}$	$p \times n$ output matrix
ϵ	small regularization parameter
$\mathcal{J}$	target functional of optimal control
$\boldsymbol{\lambda}(t)$	n-dimensional vector of time-dependent co-states for optimal control
$\boldsymbol{\mathcal{S}}$	$n \times n$ symmetric matrix of weighting coefficients
$\boldsymbol{\mathcal{P}}_{\boldsymbol{\mathcal{S}}}$	$n \times n$ projector
$\boldsymbol{\mathcal{Q}}_{\boldsymbol{\mathcal{S}}}$	$n \times n$ projector
$\boldsymbol{x}_O(t)$	state vector for outer equations
τ_L	rescaled time for left inner equations
τ_R	rescaled time for right inner equations
$\boldsymbol{X}_L(\tau_L)$	rescaled state vector for left inner equations
$\boldsymbol{X}_R(\tau_R)$	rescaled state vector for right inner equations