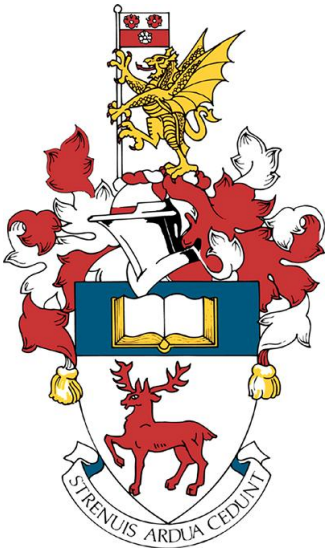
UNIVERSITY OF SOUTHAMPTON

FACULTY OF PHYSICAL AND APPLIED SCIENCES

Electronics and Computer Science



A duality-based approach to identification of linear time-varying systems

by

Joe A. Roman-Flores

A thesis submitted for the degree of MPhil

Supervisor: Prof. Paolo Rapisarda

Co-supervisor: Prof. Christopher Freeman

December 9, 2019

UNIVERSITY OF SOUTHAMPTON

ABSTRACT

FACULTY OF PHYSICAL AND APPLIED SCIENCES
Electronics and Computer Science

A thesis submitted for the degree of MPhil

A DUALITY-BASED APPROACH TO IDENTIFICATION OF LINEAR TIME-VARYING SYSTEMS

by Joe A. Roman-Flores

In this report a novel approach for the identification of linear time-varying systems is presented. We exploit the fact that external structures at the level of the inputs and outputs are reflected in the internal ones at the level of the state. Our approach first computes state trajectories from matrices of input-output data. A novel factorisation of the state trajectories from the input-output data matrices is developed. Then a state space representation compatible with the data is computed. We do not impose conditions in the time variation properties of the to-be-identified system.

A procedure for the identification of self-adjoint systems is developed. We exploit the fact that linear time-varying systems as well as nonlinear systems are self-adjoint if they have an internal representation as a linear Hamiltonian.

Finally, we utilise reproducing kernel Hilbert spaces to formalise the building of time functions for data and embed it into the duality-based approach.

Contents

Declaration of Authorship

I, Joe A. Roman-Flores , declare that the thesis entitled *A duality-based approach to identification of linear time-varying systems* and the work presented in the thesis are both my own, and have been generated by me as the result of my own original research. I confirm that:

- this work was done wholly or mainly while in candidature for a research degree at this University;
- where any part of this thesis has previously been submitted for a degree or any other qualification at this University or any other institution, this has been clearly stated;
- where I have consulted the published work of others, this is always clearly attributed;
- where I have quoted from the work of others, the source is always given. With the exception of such quotations, this thesis is entirely my own work;
- I have acknowledged all main sources of help;
- where the thesis is based on work done by myself jointly with others, I have made clear exactly what was done by others and what I have contributed myself;
- none of this work has been published before submission

Signed:.....

Date:.....

Acknowledgements

I wish to thank those people who have supported me in this journey to complete my studies:

First, I would like to express my gratitude to my supervisor Prof. Paolo Rapisarda for his valuable and constructive guidance during my studies. I would also like to thank Prof. Christopher Freeman for his support, encouragement and useful comments. To Dr. Hua-Liang Wei for accepting to examine this thesis, and Dr. Zhan Shu for accepting to be the internal examiner.

To my beloved wife Vivien Jonas for her love and understanding during the good and hard times, and to my daughter Blanka Roman Jonas for her unconditional love. It would have been impossible to do it without you.

To my late mother Blanca Flores Montoya who always inspired me to be a better person, and to my father Jose Roman Arellano who taught me that everything is possible. To all my family for their unconditional love, moral support and comfort.

To all my friends and colleagues that helped me to make this journey enjoyable. In particular, to all the VLCCC members who became not only colleagues but also very good friends during this important stage of my career. I will miss you all.

Finally, I would like to thank to my sponsors CONACYT for the scholarship, and the University of Southampton for granting me with the ECS School Studentship award.

Dedicated to the memory of my mother.
-"Juntos lo vamos a lograr" -Blanca Estela Flores Montoya

Nomenclature and notation

I_n	Identity matrix of dimension $n \times n$.
$0_{n\times m}$	Zero matrix of dimension $n \times m$.
$A^\top$	Transpose of a matrix A.
$\mathbb{R}$	The set of real numbers.
$\mathbb{N}$	The set of natural numbers.
$\mathbb{R}^w$	The space of real vectors with w components.
$\mathbb{R}^{m\times n}$	The space of $m \times n$ dimensional real matrices.
$\mathbb{R}^{\bullet\times n}$	The space of real matrices with finite unspecified number of rows and n columns.
$\mathcal{A}$	The space of analytic functions from $\mathbb{R}$ to $\mathbb{R}$.
rank (A)	Rank of the matrix A.
$\text{col}(A, B)$	$\text{col}(A, B) := [A^\top \ B^\top]^\top$.
$\Pi_B(A)$	The projection of the row space of the matrix A onto the row space of the matrix B.
$A^\dagger$	Moore−Penrose pseudoinverse of the matrix A.
$(S\backslash V)$	Set difference.

Chapter 1

Introduction

1.1 System identification

System identification is a field that provides mathematical tools to construct models of dynamical systems from measurements. These models obtained from data are not the exact model, rather they are models which are compatible with the given data. Thus, system identification is meaningful in the industrial field, since models for industrial processes are often difficult or impossible to obtain from physical principles due to a high level of complexity.

The need for accurate models of engineering applications is encouraging the development of new identification methods. In recent decades, system identification has been widely exploited, particularly for *linear time-invariant* (LTI) systems. As a result, the collection of available methods has become vast for LTI systems. However, systems in practice are often of a *nonlinear* or *linear time-varying* (LTV) nature requiring further study of system identification for such areas.

1.1.1 LTV identification

LTV system identification approaches can be classified as either: using one input-output trajectory and fitting the data to a structure determined by *a priori* assumptions on the system order and the nature (polynomial, periodic, etc.) of its nonstationarity (see e.g. Li et al. (2011)), or those working on an *ensemble* of data that is assumed to share common time-varying features, see e.g. Ba et al. (2016). Such data can be of a special nature, e.g. impulse-response data or general input-output trajectories.

Methods based on fitting input-output data to a prior determined structure need first to analyse the data and then select a model class that will best describe the behaviour of the system. The choice of a particular system structure for the system to be identified

is based on the application and the physical nature of the system. Then time-varying coefficients are identified turning the procedure to a type of parametric identification. These methods usually capture the dynamics of the system in predetermined time-varying functions used as parameters.

For methods based on an ensemble of data, subspace identification provides the attractive state space model-based approach. It may seem that the class of system is restricted, however, many physical systems are accurately modelled in the state space approach. Moreover, tools and methodologies for simulation and control in this class of systems are well established. Subspace identification methods have become a popular choice for the identification of LTV models due to the use of state space models which also provides advantages in multi-input-multi-output system identification. However, for current methods the extraction of subspaces of data matrices is based on priori knowledge of time-varying properties.

1.1.2 Duality for system identification

The duality approach for system identification has been developed in recent years underlying several previous results in identification and model-order reduction of linear time-invariant systems, see Rapisarda and Rao (2013); Rapisarda and Antoulas (2015a,b); Rapisarda and der Schaft (2013); Rapisarda and Trentelman (2011). Duality exploits the fact that external properties at the level of the inputs and outputs are reflected in the internal ones at the level of the states. This connection is given by a bilinear form of the external variables. Thus, the starting point of the identification approach is data gathered from the unknown system.

Similar to subspace identification, the duality approach exploits the state space model class. Then, the advantages of this class of systems are also presented in this approach. Furthermore, no prior knowledge of the to be identified system is required, e.g. the choice of the system structure. This approach has shown to be a power tool for system identification for LTV as we will show in this thesis.

1.2 Motivation

LTV models play a significant role in many research areas, such as biology, econometrics, and engineering. Robotic mechanisms, aerospace structures and bridges are examples of structures which exhibit time-varying properties due to vibration or physical deformation. Identification methods also become an important requirement for applications in control and simulation.

On one hand, the complexity of systems in engineering applications is increasing, and the modelling of such systems is becoming a big challenge. For example, the interconnection

of multiple systems, e.g. switched systems in power systems, increase the complexity of the modelling. On the other hand, system identification provides useful tools to get accurate models to achieve important goals in application such as simulation, analysis and control.

Current methods for LTV system identification are based on prior knowledge of the time-variation properties of the system to be identified. This thesis develops a new identification approach for LTV systems using duality. We will exploit the connection of the external variables with the state variables given by duality to compute a state space representation compatible with the data. The duality-based approach provides useful conditions and tools to allow the identification of LTV systems without prior assumption of time-variation properties using multiple system responses.

1.3 Outline of this thesis

This thesis is presented in the following order:

- *Chapter 2*: An overview of time-varying system identification in the actual literature is given. We review LTV identification algorithms and their crucial steps for the identification process.
- *Chapter 3*: The notion of time-varying systems in the state space approach is introduced. We present the concept of equivalent transformations. Relevant concepts of controllability and observability are provided. We also introduce *reproducing kernel Hilbert spaces*.
- *Chapter 4*: Duality for linear time-varying system is defined. The self-adjointness case is introduced.
- *Chapter 5*: In this chapter we develop the duality-based approach for LTV system identification. We develop algorithms for the self- and nonself-adjoint case.
- *Chapter 6*: We implement the framework developed in this thesis to Hamiltonian systems.
- *Chapter 7*: We apply the duality based approach to a real-time system. Simulation and experimental results are provided.

Finally we provide an Appendix A containing definitions needed for this thesis.

Chapter 2

An overview of LTV system identification

In this chapter, we discuss some identification techniques for linear time-varying systems available in the literature. We outline and discuss the main steps for existing system identification algorithms. Finally, a motivation for the proposed framework is given.

2.1 System identification for LTV systems

The general problem of system identification is to determine the most suitable model for describing the data generated by the to-be-identified system.

In order to obtain a model using system identification theory, we have to consider several critical stages known as the identification cycle, see Fig. 2.1. These steps are based on several identification procedures in the literature (see e.g. Verhaegen and Verdult (2007)), and explained in the following.

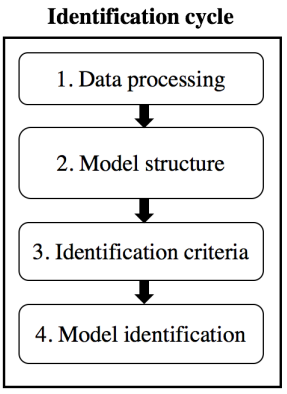


Figure 2.1: Identification cycle.

1. Data processing: This stage refers to the experimental design that allows gathering the required data from the to-be-identified system. The nature of these data will depend on the physical nature of the system, e.g. voltage and current in an electrical circuit or position and velocity in a mass-spring system.

2. Model structure: One of the most crucial step in the identification cycle is the choice of a model structure. A model structure is chosen from a set of candidates models to fit the data. This choice is made considering the type of representation for the data, e.g. state space representation, input-output representation, etc.

3. Identification methodologies: In this part, the user decides the processes and methodologies to achieve a correct identification procedure. This can include extraction of a state sequence corresponding to a state space representation from data, or the computation of parameters in a input-output model.

4. Model identification: The model identification stage is the application of the tools and methodologies described previously.

In the following we address the current identification methodologies for LTV system identification and discuss their procedures. We will see later how the different model structures and identification methodologies influence the results of the identification procedure.

2.2 LTV input-output model based approach

In the existing literature, several methods have adopted input-output-based model as a model structure for time-varying system identification. Particularly, a filter-based approach has been used to provide an algorithm using iterative processes to track time-varying parameters fitting a chosen model. This approach follows the identification cycle described in Fig. 2.1 under certain assumptions about the time variation of the system parameters. The identification procedure is based on the a priori knowledge of the parameter time variation, e.g. fast or slow time variation. In this section, we introduce the least-squares problem which is the basis of these approaches. We explain the two main approaches used for LTV system identification using a filter-based approach.

2.2.1 A least-squares approach

The least-squares is a standard approach in regression methods to solve overdetermined set of equations, see Verhaegen and Verdult (2007). Least-squares has been widely exploited in system identification for linear systems and adapted for linear time-varying systems. In the following we introduced some basis of the general least-squares problem.

Consider a data matrix $X \in \mathbb{R}^{m\times n}$ whose $m,n-$th element is the $m-$th observation of the $n-$th independent variable with $m > n$, , an observation vector $y \in \mathbb{R}^m$ and a vector of unknown parameters $\theta \in \mathbb{R}^n$. The matrix equation in the unknown parameter vector is given by

$$y = X\theta. \tag{2.1}$$

The above equation can be solved for θ only if y lies in the column space of X. If this is not the case, to find a solution we consider the equation

$$y = X\theta + e, \tag{2.2}$$

where e is called the residual vector (or error). Provided that the matrix X has full column rank, the ordinary least squares (OLS) approach is used to solve (2.2). The OLS is then formulated as an optimization problem of finding a vector θ that minimizes the norm of the residual vector e s.t. $y \approx X\theta$, i.e.

$$\min_{\theta} \parallel X\theta - y \parallel_2^2 . \tag{2.3}$$

The cost function of the least-squares problem can be expanded as follows

$$\begin{aligned} f(\theta) = &\parallel X\theta - y \parallel_2^2 \\ =&(X\theta - y)^\top (X\theta - y) \\ =&(\theta^\top X^\top - y^\top)(X\theta - y) \\ =&\theta^\top X^\top X\theta - y^\top X\theta - \theta^\top X^\top y + y^\top y. \end{aligned} \tag{2.4}$$

The solution to the least-squares problem is given by computing the gradient $\partial f(\theta)/\partial\theta$ and setting it to zero (see Verhaegen and Verdult (2007)), i.e.

$$\frac{\partial f(\theta)}{\partial \theta} = 2X^\top X\theta - 2X^\top y = 0. \tag{2.5}$$

Then the solution for (2.3) can be express as

$$X^\top X\theta - X^\top y = 0. \tag{2.6}$$

From the assumption that the matrix X has full column rank, we can compute $\widehat{\theta}$ that minimizes (2.3) by multiplying (2.6) on the left by $(X^\top X)^{-1}$, then

$$\widehat{\theta} = (X^\top X)^{-1} X^\top y. \tag{2.7}$$

If the rank of X is not full column rank, the matrix $(X^\top X)$ cannot be inverted, therefore the solution in (2.7) does not exist. In this case, one has to use a rank-deficient least squares approach, see for example Santos et al. (1998).

2.2.2 Adaptive filter algorithm for system identification

We now introduce an adaptive filter algorithm based on the least-squares problem for system identification. Let an input signal be denoted by the sample set $u(t), u(t-1), .., u(t-T+1)$ where $u(t)$ and $u(t-1)$ denotes the input signal at time t and $t-1$, respectively, and T is the number of samples to be considered. This input is applied to a LTV system (the to-be-identified system), and the corresponding output signal $y(\cdot)$ is generated. Consider the following single-input-single-output linear relationship

$$y(t) = \sum_{i=1}^{m} a_i(t)y(t-i) + \sum_{j=1}^{p} b_j(t)u(t-j) + v(t) \tag{2.8}$$

where y and u are the sampled output and input of the to-be-identified system, v is the error signal, a_i and b_j are the time-varying unknown parameters to be determined, m and p are the maximum number of parameters for y and u, respectively. The value of m and p is determine by the number of samples considered.

Define

$$\varphi(t) := \begin{bmatrix} y(t-1) & \cdots & y(t-m) & u(t-1) & \cdots & u(t-p) \end{bmatrix}^\top$$

and

$$\theta(t) := \begin{bmatrix} a_1(t) & \cdots & a_m(t) & b_1(t) & \cdots & b_p(t) \end{bmatrix}^\top$$

then, the model in 2.8 can be rewritten as

$$y(t) = \varphi(t)^\top \theta(t) + v(t). \tag{2.9}$$

It has been shown that least squares-based algorithms can be used to compute the time-varying parameters in (2.9) under the assumption that the time-variation is relatively slow, see e.g. Isermann and Münchhof (2014) Sec. 9.6 and Niedzwiecki and Klaput (2002). This assumption allows to perform the identification procedure in local windows where the system behaves as time-invariant, see Fig. 2.2.

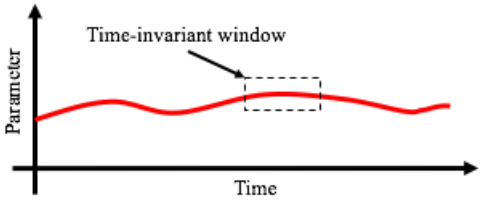


Figure 2.2: Slow time-varying parameters with time-invariant window.

Particularly, *exponentially weighted least square* (EWLS) has been used for LTV system identification. Similar to the classical approach of least squares, EWLS estimates the parameters of the model by minimizing the sum of the squared error between the

observed responses and the output of the proposed model. However, EWLS includes exponentially decaying weights to define the influence of the past data in the parameters to be estimated, i.e. to compute the parameters one has to solve the following minimization problem:

$$\overline{\theta}(t) = \arg\min_{\theta} \sum_{i=0}^{t-1} w(i) \left[y(t-i) - \varphi(t-i)^{\top} \theta \right]^2 \tag{2.10}$$

where $w(i) = \sigma^i$ is the exponentially decaying weight. The choice of σ depends on the user, and it is based on desired objective, e.g. for suppression of disturbances σ has to be chosen closer to 1 and for better time-varying tracking $\sigma < 1$. Note that the parameters computed with EWLS are denoted by $\overline{\theta}$ to emphasize that they correspond to an approximation of the true parameters.

The solution for (2.10) is given by the following expression

$$\overline{\theta}(t) = \left[\textstyle\sum_{i=0}^{t-1} \sigma^i \varphi(t-i)\varphi(t-i)^{\top} \right]^{-1} \left[\textstyle\sum_{i=0}^{t-1} \sigma^i y(t-i)\varphi(t-i) \right]. \tag{2.11}$$

For online identification, a recursive algorithm is adopted. Consider (2.9), then the recursive estimator takes the form

$$\overline{\theta}(t) = \overline{\theta}(t-1) + L(t) \left[y(t) - \varphi(t)^{\top} \theta(t-1) \right] \tag{2.12}$$

where the new estimation $\overline{\theta}(t)$ depends on the previous estimation $\overline{\theta}(t-1)$ and a weighted prediction error with $L(t)$ being a weighting matrix. The process of how to choose $L(t)$ as been discussed in Keesman (2011) and it is given by

$$\begin{aligned} L(t) &= \frac{P(t-1)\varphi(t)}{\sigma + \varphi(t)^{\top} P(t-1)\varphi(t)} \\ P(t) &= \sigma^{-1}[P(t-1) - L(t)\varphi(t)^{\top} P(t-1)] \end{aligned} \tag{2.13}$$

where $P(t)$ is a covariance matrix. The well-known recursive algorithm for parameter computation is given in algorithm 1, see Niedzwiecki and Klaput (2002). The recursive algorithm 1 is also shown in Fig. 2.3.

The assumptions for the the EWLS algorithms are:

1. The behaviour of the time varying properties is known, i.e. slow time varying.
2. The time variations are slow so that the time-varying process is locally time invariant.
3. Suitable weight values are known for an efficient identification process.

Algorithm 1 EWLS algorithm

Input: $(y(t-i), u(t-j))_{i=1,..,m\ j=1,..,p}$
Output: $\overline{\theta}(t)$ s.t. $\varphi(t)^\top \overline{\theta}(t) + e(t) \approx \varphi(t)^\top \theta(t) + v(t)$
Assumptions: Slowly time-varying, suitable weights σ known.

1. Initialization at $t = 0$:

$$\begin{aligned}
&\overline{\theta}(0) = 0\\
&P(0) = \mathrm{diag}(\alpha_1, .., \alpha_{m+p}) \text{ (with } \alpha_\mathrm{i} \text{ being large constants)}\\
&\varphi(t) = \begin{bmatrix} y(t-1) & \cdots & y(t-m) & u(t-1) & \cdots & u(t-p) \end{bmatrix}^\top
\end{aligned}$$

2. Store new measurements in $(y(t), u(t))$ then compute

$$e(t) = y(t) - \varphi(t)^\top \overline{\theta}(t-1)$$

and

$$L(t) = \frac{P(t-1)\varphi(t)}{\sigma + \varphi(t)^\top P(t-1)\varphi(t)}.$$

3. New parameter estimation

$$\overline{\theta}(t) = \overline{\theta}(t-1) + L(t)e(t)$$

4. Compute

$$P(t) = \sigma^{-1}[P(t-1) - L(t)\varphi(t)^\top P(t-1)]$$

5. Set $t = t + 1$, and start again from step 2.

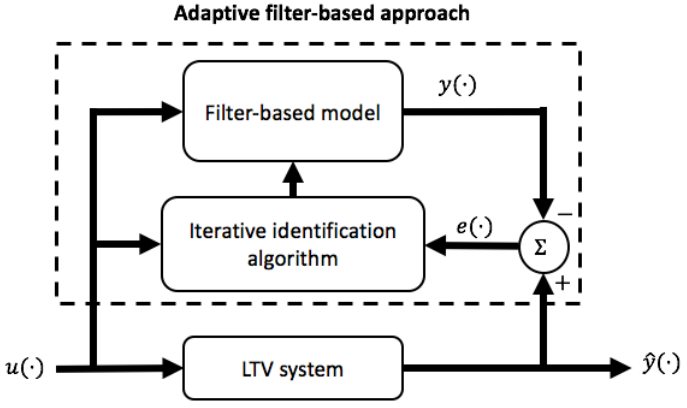


Figure 2.3: Adaptive filter-based identification

A disadvantage of weighted least squared-based algorithms is that, in the choice of the corresponding weights, one has to choose between an accurate estimation of the time-varying parameters or a robust identification process. Moreover, the prior knowledge of the time variation properties becomes cumbersome in system identification if one does

not have enough information of the system to be identified.

2.2.3 Basis function-based method for system identification

To identify systems of more complexity or fast time-varying processes, it was shown that the system parameters can be represented using a linear combination of predetermined functions, also known as basis functions. These functions are chosen from a set of candidate basis functions (they can be of different nature, e.g exponential as in de Almeida Rego et al. (2014) or polynomial as in Huang et al. (2009)) depending on the signal behaviour.

Consider the system described by eq. (2.8), and let B be a set of k linearly independent basis functions, i.e.

$$B := \{c_1(\cdot), c_2(\cdot), .., c_k(\cdot)\},$$

then the time-varying parameters are approximated by a linear combination of the preselected basis functions, i.e.

$$\begin{aligned} a_i(t) &= \sum_{j=1}^{k} \alpha_{ij} c_j(t) \\ b_i(t) &= \sum_{j=1}^{k} \beta_{ij} c_j(t) \end{aligned} \tag{2.14}$$

where k is maximum number of basis functions, α_{ij} and β_{ij} are constant parameters to be determine. Now we can rewrite (2.8) using (2.14) as

$$y(t) = \sum_{i=1}^{m} \sum_{j=1}^{k} \alpha_{ij} c_j(t) y(t-i) + \sum_{i=1}^{p} \sum_{j=1}^{k} \beta_{ij} c_j(t) u(t-i) + v(t), \tag{2.15}$$

then (2.15) can be rewritten as

$$y(t) = \sum_{j=1}^{k} c_j(t) \begin{bmatrix} y(t-1) & \cdots & y(t-m) & u(t-1) & \cdots & u(t-p) \end{bmatrix} \begin{bmatrix} \alpha_{1,j} \\ \vdots \\ \alpha_{m,j} \\ \beta_{1,j} \\ \vdots \\ \beta_{p,j} \end{bmatrix} + v(t)$$

or compactly as

$$y(t) = \sum_{j=1}^{k} c_j(t) \varphi(t)^\top \Theta_j + v(t) \tag{2.16}$$

where

$$\varphi(t) := \begin{bmatrix} y(t-1) & \cdots & y(t-m) & u(t-1) & \cdots & u(t-p) \end{bmatrix}^\top$$

and

$$\Theta_j := \begin{bmatrix} \alpha_{1,j} & \cdots & \alpha_{m,j} & \beta_{1,j} & \cdots & \beta_{p,j} \end{bmatrix}^\top.$$

In this approach the preselected basis functions are used to describe the time-varying behaviour of the parameters. Thus, we can now treat (2.16) as time-invariant, since the unknown parameters to be computed are those related to the basis functions, i.e. α_j and β_j, which are time-invariant parameters. The procedure to solve eq. (2.16) follows the classical least-square approach as in Sec. 2.2.2, this is due to the time-invariant form, see Li et al. (2016).

The critical stage in this approach is the selection of appropriate basis functions that will lead to an accurate estimation of the time-varying parameters.

The resulting least squares-based approach using basis function assumes:

1. Prior knowledge of the system to be identified for the selection of appropriate basis functions,
2. A suitable choice of the maximum number of basis functions to be used.

When no prior knowledge of the system is available, a blind choice for the basis functions is made using "generic" basis functions. However, this choice is crucial since it has been shown that these methods suffers from sensitivity to the choice of pre-design parameters, e.g. weights in WLS approach.

2.3 LTV state space model based approach

State space representation is widely accepted in engineering applications due to its advantages in analysis and control. An important concept in the state space approach is the definition of the *state*.

A state at a time t and given future inputs are the sufficient information needed to determine the future evolution of the state. In other words, the state at time t has sufficient information from the past to determine the future evolution, see Dewilde and der Veen (1998).

From this point of view, several system identification methods based on state space representation take advantage of this property as we shall see later. In the following, we provide an overview of identification methods in the literature based on state space approach.

2.3.1 The lower triangular matrix case

Several methods use the discrete version of the state space representation. A minimal representation in a discrete form is given by:

$$\begin{aligned} x(t+1) &= A(t)x(t) + B(t)u(t) \\ y(t) &= C(t)x(t) + D(t)u(t) \end{aligned} \tag{2.17}$$

where x is the state variable of dimension n, u the input variable of dimension m, y the output variable of dimension p, and matrices A, B, C and D of appropriate dimension.

To define the identification problem, first we analyse the equations in (2.17) assuming that there exists already a given representation for the to-be-identified system. Then we will find the conditions for identification. Assume that the initial state is zero, i.e. $x(0) = 0$, then it follows that for $t = 0$

$$\begin{aligned} x(1) &= B(0)u(0) \\ y(0) &= D(0)u(0), \end{aligned}$$

and for $t = 1$ we have that (2.17) now becomes

$$\begin{aligned} x(2) &= A(1)x(1) + B(1)u(1) \\ y(1) &= C(1)x(1) + D(1)u(1), \end{aligned}$$

but since $x(1) = B(0)u(0)$ it follows that

$$\begin{aligned} x(2) &= A(1)B(0)u(0) + B(1)u(1) \\ y(1) &= C(1)B(0)u(0) + D(1)u(1). \end{aligned}$$

Assume that the impulse response of (2.17) is known. Now we can express an input-output map by eliminating the state as above for $t = 0, .., N$ as follows:

$$\begin{bmatrix} y(0) \\ y(1) \\ y(2) \\ \vdots \\ y(N) \end{bmatrix} = \begin{bmatrix} D(0) & 0 & \cdots & 0 \\ C(1)B(0) & D(1) & 0 & 0 \\ C(2)A(1)B(0) & C(2)B(1) & D(2) & 0 \\ \vdots & \vdots & \vdots & \vdots \\ C(N)A(N-1)\cdots A(1)B(0) & \cdots & C(N)B(N-1) & D(N) \end{bmatrix} \begin{bmatrix} u(0) \\ u(1) \\ u(2) \\ \vdots \\ u(N) \end{bmatrix}. \tag{2.18}$$

As it is evident from the above expression, the construction of the central matrix in (2.18) is given by projecting (2.17) to future inputs. Now we express (2.18) in a simpler

manner by denoting the central matrix by M and

$$\begin{aligned} U :=& \text{col}(u(0), u(1), .., u(N)) \\ Y :=& \text{col}(y(0), y(1), .., y(N)), \end{aligned}$$

then (2.18) in a simpler form becomes $Y = MU$. Note that M is lower block triangular and its non-zero blocks are submatrices of dimension $p \times m$. We can compact the notation of the submatrices of M to simplify their analysis. We then denote $T_{i,j}$ as a submatrix of dimension $p \times m$ defined by

$$\begin{aligned} T_{i,j} &:= D(i) \quad \text{for } i = j, \\ T_{i,j} &:= C(i)A(i-1)A(i-2)\cdots A(j+1)B(j) \quad \text{for } i - 1 > j \\ T_{i,j} &:= C(i)B(j) \quad \text{for } i - 1 = j \\ T_{i,j} &:= 0 \quad \text{for } i < j, \end{aligned} \tag{2.19}$$

e.g. $T_{0,0}$ denotes the top-left block submatrix of M in (2.18), i.e. $T_{0,0} := D(0)$, and in a similar way $T_{1,0} := C(1)B(0)$ and $T_{2,0} := C(2)A(1)B(0)$. It follows that M in (2.18) is given by

$$M = \begin{bmatrix} T_{0,0} & 0 & \cdots & 0 \\ T_{1,0} & T_{1,1} & 0 & 0 \\ T_{2,0} & T_{2,1} & T_{2,2} & 0 \\ \vdots & \vdots & \vdots & \vdots \\ T_{N,0} & \cdots & T_{N,N-1} & T_{N,N} \end{bmatrix}.$$

Now consider that M is given without structure, i.e. a constant matrix which maps $u \mapsto y$. This is equivalent to say that the time-varying impulse response of the system is known. Thus the identification problem consists in extracting the information of suitable system matrices $A(\cdot), B(\cdot), C(\cdot), D(\cdot)$ of a state space realization from M. The identification problem is known as the realization problem (see e.g. Dewilde and der Veen (1998)), and the conditions for identification are as follows.

Consider again the input-output equation $Y = MU$. Then define the input sequence as $U := \{u(\ell)\}_{\ell=-\infty}^{\infty}$. Considering k a reference index, we can divide this input sequence into past inputs U_{p_k} and future inputs U_{f_k}, where U_{p_k} defines the input up to $k-1$, and U_{f_k} the input from k up to infinity. i.e.

$$\begin{aligned} U_{p_k} :=& \text{col}(.., u(k-2), u(k-1)) \\ U_{f_k} :=& \text{col}(u(k), u(k+1), ..). \end{aligned}$$

Similarly, for an output sequence $Y = \{y(\ell)\}_{\ell=-\infty}^{\infty}$ we define

$$\begin{aligned} Y_{p_k} &:= \text{col}(.., y(k-2), y(k-1)) \\ Y_{f_k} &:= \text{col}(y(k), y(k+1), ..). \end{aligned}$$

Then we can rewrite the input-output equation with the past and future data as

$$\begin{bmatrix} Y_{p_k} \\ Y_{f_k} \end{bmatrix} = \begin{bmatrix} M_{p_k} & 0 \\ H_k & M_{f_k} \end{bmatrix} \begin{bmatrix} U_{p_k} \\ U_{f_k} \end{bmatrix}. \tag{2.20}$$

with matrices M_{p_k}, H_k and M_{f_k} of appropriate dimensions.

Now, looking at (2.20) we know that H_k has the following structure:

$H_k :=$

$$\begin{bmatrix} \cdots & C(k)A(k-1)A(k-2)B(k-3) & C(k)A(k-1)B(k-2) & C(k)B(k-1) \\ \cdots & C(k)A(k)A(k-1)A(k-2)B(k-3) & C(k+1)A(k)A(k-1)B(k-2) & C(k+1)A(k)B(k-1) \\ \cdots & C(k+2)A(k+1)A(k)A(k-1)A(k-2)B(k-3) & C(k+2)A(k+1)A(k)A(k-1)B(k-2) & C(k+2)A(k+1)A(k)B(k-1) \\ \vdots & \vdots & \vdots & \vdots \end{bmatrix}$$

which can be factorized as

$$H_k = \begin{bmatrix} C(k) \\ C(k+1)A(k) \\ C(k+2)A(k+1)A(k) \\ \vdots \end{bmatrix} \begin{bmatrix} \cdots & A(k-1)A(k-2)B(k-3) & A(k-1)B(k-2) & B(k-1) \end{bmatrix}. \tag{2.21}$$

Factorizations of H_k can be computed using SVD as in Van Der Veen et al. (1993). In (2.21), we can identify what is known as the observability operator at index k

$$\mathrm{O}_k := \begin{bmatrix} C(k) \\ C(k+1)A(k) \\ C(k+2)A(k+1)A(k) \\ \vdots \end{bmatrix}$$

and the controllability operator at index k:

$$\mathrm{C}_k := \begin{bmatrix} \cdots & A(k-1)A(k-2)B(k-3) & A(k-1)B(k-2) & B(k-1) \end{bmatrix}.$$

Assume that H_k is defined in terms of a minimal realization. From the factorization in (2.21), it follows that the rank of H_k is less than or equal to the numbers of columns of O_k and rows of C_k, e.g. if A is an n by n matrix, then $\text{rank}(H_k) = n$. It is said that the realization corresponding to H_k is minimal if the rank of H_k is n for all k.

To obtain the system matrices we proceed as follows. From (2.19) $D(k) = T_{k,k}$, and $C(k)$ and $B(k)$ can be extracted from a minimal factorization of H_k as from (2.21)

$$C(k) = [\text{ First } p \text{ rows of } \mathrm{O}_k]$$
$$B(k) = [\text{ Last } m \text{ columns of } \mathrm{C}_{k+1}]$$

For obtaining $A(k)$, we know that the columns of O_k are linearly independent for all k if the realization is minimal. Then O_k has a left inverse, i.e. there exists N_k s.t. $N_k\mathrm{O}_k = I$ for all k. Then we can compute $A(k)$ as

$$A(k) = N_{k+1}\mathrm{O}_k.$$

Note that the choice of the matrix inverses to compute the system matrices will define the basis of the state space.

The identification algorithm for the matrix case is then given in algorithm 2.

Algorithm 2 Matrix case algorithm

Input: M lower triangular matrix in terms of data, i.e. time-varying impulse response
Output: $A(\cdot), B(\cdot), C(\cdot), D(\cdot)$, s.t. (2.17) holds
Assumptions: Available impulse response data

1. Set $k = 0$
2. Compute the rank of H_k, i.e. $\text{rank}(H_k) = n$
3. Compute a minimal factorization $H_k = \mathrm{O}_k\mathrm{C}_k$ as in (2.21)
4. Compute:
   - $A(k) = N_{k+1}\mathrm{O}_k$ s.t. $N_k\mathrm{O}_k = I$
   - $B(k) = [\text{first } p \text{ rows of } \mathrm{O}_k]$
   - $C(k) = [\text{first } m \text{ columns of } \mathrm{C}_{k+1}]$
   - $D(k) = T_{k,k}$
5. Set $k = k + 1$, and start from Step 2

The assumption that plays an important role in this identification approach is the accessibility of the entries of a matrix that produces the map $u \mapsto y$, i.e. the availability of the impulse responses of the system to be identified. However, in a more general case, one does not have access to such information. Thus, methods for identification using observed input-output measurements are more suitable when little prior information of the system is known. This is discussed in the following subsection.

2.3.2 Subspace identification

Subspace-based identification methods are similar to the lower triangular matrix case, in the sense that the state space structure is considered. However, the ideas of subspace identification lie in the fact subspaces can be extracted from Hankel matrices of input and output data, and such subspaces are related to the system matrices of a state space representation for the data. Subspace algorithms use geometric tools and numerical linear algebra to conduct such procedures, e.g. projection, *QR decomposition* and *singular value decomposition* (SVD).

Several subspaces-based approaches have been developed for the LTV case over the years, see Verhaegen and Yu (1995), Shi et al. (2007), Ohsumi and Kawano (2002), Jhinaoui et al. (2012). Particularly in the subspace-based identification approach for LTV system, Hankel matrices of time-varying processes are used to extract the observability matrix of a state space model for the data. We can find identification algorithms that use an ensemble of input-output data of a single response as in Shi et al. (2007). This single responses matrix are defined in a Hankel form, i.e. the Hankel matrix associated with an output $y(k)$ $k = 1, .., N$ is defined as

$$H(y(k))_{h_1,h_2} := \begin{bmatrix} y(k) & y(k+1) & \cdots & y(k+h_2-1) \\ y(k+1) & y(k+2) & \cdots & y(k+h_2) \\ \vdots & \vdots & \cdots & \vdots \\ y(k+h_1-1) & y(k+h_1) & \cdots & y(k+h_1+h_2-2) \end{bmatrix} \tag{2.22}$$

where the index h_1 and h_2 indicate the number of rows and columns, respectively, and they are appropriately chosen such that the identification procedure is guaranteed. The index k indicates the time instant of the upper left element of the matrix. In the case when $y(k) \in \mathbb{R}^p$, the number of rows and columns is $h_1 p$ and h_2, respectively. A Hankel matrix $H(u(k))_{h_1,h_2}$ for the input $u(k)$ $k = 1, .., N$ is built in a similar way.

As we can see in (2.22), the number of rows and columns grows as the time interval of the response increases. This leads to large matrices of data when large time-intervals are needed for analysis increasing the computational complexity in applications. The choice of h_1 and h_2 are chosen so that they guarantee the observability and controllability of the system.

Consider the matrix in (2.22). Since the indexes h_1 and h_2 are fixed, and do not change during the identification procedure, we simplify the notation by removing them, i.e. we now refer to (2.22) as $H(y(k))$, and $H(u(k))$ for $u(k)$. From (2.22) and (2.17), it follows that

$$H(y(k)) = \mathcal{O}(k)X_d(k) + \mathcal{T}(k)H(u(k)). \tag{2.23}$$

where $X_d(k) := \text{diag}(x(k), x(k+1), \cdots, x(k+h_2-1))$, and

$$\mathcal{O}(k) := \begin{bmatrix} C(k) & \cdots & C(k+h_2-1) \\ C(k+1)A(k) & \cdots & C(k+h_2)A(k+h_2-1) \\ \vdots & \cdots & \vdots \\ C(k+h_1-1)A(k+h_1-2)\cdots A(k) & \cdots & C(k+h_1+h_2-2)\cdots A(k+h_2-1) \end{bmatrix}$$

$$\mathcal{T}(k) := \begin{bmatrix} D(k) & 0 & 0 & 0 & 0 \\ C(k+1)B(k) & D(k+1) & 0 & 0 & 0 \\ \vdots & \ddots & \ddots & 0 & 0 \\ \vdots & \vdots & \ddots & \ddots & 0 \\ C(k+h_1-1)A(k+h_1-2)\cdots A(k+1)B(k) & \cdots & \cdots & \cdots & D(k+h_1-1) \end{bmatrix}$$

Since the matrices $\mathcal{O}(k)$ and $\mathcal{T}(k)$ contains the information of the system matrices for (2.17), they play an important role in the identification procedure. $\mathcal{O}(k)$ is called the observability matrix and $\mathcal{T}(k)$ the controllability matrix. The system (2.17) is said to be observable if $\mathcal{O}(k)$ has full column rank. In a like manner, (2.17) is said to be controllable if $\mathcal{T}(k)$ has full row rank.

Based on several identification methods for LTV system (see e.g. Shi et al. (2007) and Liu (1997)), it has been shown that under the assumption that the system is slowly time-varying, then the matrix $\mathcal{O}(k)$ can be approximated as

$$\mathcal{O}(k) \approx \begin{bmatrix} C(k) \\ C(k+1)A(k) \\ \vdots \\ C(k+h_1-1)A(k+h_1-2)\cdots A(k) \end{bmatrix}.$$

Then, considering the above expression for $\mathcal{O}(k)$, (2.23) can be written as

$$H(y(k)) = \mathcal{O}(k)X(k) + \mathcal{T}(k)H(u(k)) \tag{2.24}$$

where $X(k) := [x(k)\ x(k+1)\ \cdots\ x(k+h_2-1)]$.

Note that $H(y(k)) \in \mathbb{R}^{ph_1 \times nh_2}$, $\mathrm{X}(k) \in \mathbb{R}^{nh_2 \times h_2}$, $\mathcal{O}(k) \in \mathbb{R}^{ph_1 \times n}$, $H(u(k)) \in \mathbb{R}^{mh_1 \times nh_2}$ and $\mathcal{T}(k) \in \mathbb{R}^{ph_2 \times mh_1}$. Here $\mathcal{O}(k)$ is called the observability matrix of (2.17) in terms of $H(y(k))$ and $H(u(k))$. An important stage in the subspace algorithm is the extraction of the column space of such observability matrix from the input and output data matrices. This is done with different methodologies. One is by isolating the desired term in (2.24) by assuming that h_1 and h_2 have been properly chosen such that we can define a full row rank matrix $H(u(k))^\perp$ orthogonal to $H(u(k))$ s.t. $H(u(k))H(u(k))^\perp = 0$, e.g.

$$H(u(k))^\perp = I - H(u(k))^\top \left[H(u(k))H(u(k))^\top\right]^{-1} H(u(k)),$$

then post multiplying (2.24) with $H(u(k))^{\perp}$ we obtain

$$H(y(k))H(u(k))^{\perp} = \mathcal{O}(k)X(k)H(u(k))^{\perp}. \tag{2.25}$$

From the above equation, it is evident that we can obtain information of the observability matrix from matrices of input-output data. This can be done using a classical SVD-based method as in Shi et al. (2007). This is possible assuming that the system is observable in the sense that $\mathcal{O}(k)$ has full column rank, then it follows that

$$\text{Column space}\left[H(y(k))H(u(k))^{\perp}\right] = \widehat{\mathcal{O}}(k).$$

Performing SVD we have that

$$H(y(k))H(u(k))^{\perp} = R(k)\Sigma(k)V(k)^{\top},$$

it follows that the first n columns of $R(k)$ provide a set of base vectors for the column space of $\widehat{\mathcal{O}}(k)$, i.e. a observability matrix for the data $H(y(k))$ and $H(u(k))$ can be defined as

$$\widehat{\mathcal{O}}(k) := [\text{ first } n \text{ columns of } R(k)]. \tag{2.26}$$

Similar procedures are performed for the controllability matrix $\mathcal{T}(k)$ assuming that the system is controllable in the sense that $\mathcal{T}(k)$ has full row rank. It follows that by defining a matrix $\widehat{\mathcal{O}}(k)^{\perp}$ orthogonal to $\widehat{\mathcal{O}}(k)$ such that $\widehat{\mathcal{O}}(k)^{\perp}\widehat{\mathcal{O}}(k) = 0$ the controllability matrix is obtained pre-multiplying (2.24) by $\widehat{\mathcal{O}}(k)^{\perp}$, i.e.

$$\widehat{\mathcal{O}}(k)^{\perp}H(y(k)) = \widehat{\mathcal{O}}(k)^{\perp}\widehat{\mathcal{T}}(k)H(u(k)),$$

then, it follows that

$$\widehat{\mathcal{T}}(k) = [\widehat{\mathcal{O}}(k)^{\perp}]^{\dagger}\widehat{\mathcal{O}}(k)^{\perp}H(y(k))[H(u(k))]^{\dagger}.$$

Finally, system matrices for the input-output data are extracted from the observability and controllability matrix. The procedure is summarised in algorithm 3.

Subspace identification methods for LTV systems of a single response need the assumption of slowly time-varying processes. This means that the time-varying properties of the system have to be known a priori. We are also aware of subspace methods using a ensemble of input-output data as in Verhaegen and Yu (1995). These approaches follows the SVD-based method as in the single response, however, their computational complexity increases as several responses are included in the analysis as well as the time-interval,

Algorithm 3 Subspace Algorithm

Input: $(u(k), y(k))$ for $k = 1, .., N$
Output: $A(k), B(k), C(k), D(k)$ for $k = 1, .., N$ s.t. (2.17) holds
Assumptions: Slowly time-varying, controllable and observable.

1. Set $k = 1$
2. Compute $H(u(k))$, $H(y(k))$
3. Compute $H(u(k))^{\perp}$ s.t. $H(u(k))H(u(k))^{\perp} = 0$
4. Perform SVD:
$$H(y(k))H(u(k))^{\perp} = R(k)\Sigma(k)V(k)^{\top}$$
$$H(y(k+1))H(u(k+1))^{\perp} = R(k+1)\Sigma(k+1)V(k+1)^{\top}$$
5. Define
$$\widehat{\mathcal{O}}(k) := [\text{ first } n \text{ columns of } R(k)]$$
$$\widehat{\mathcal{O}}(k+1) := [\text{ first } n \text{ columns of } R(k+1)]$$
6. Compute $\widehat{\mathcal{O}}(k)^{\perp}$ s.t. $\widehat{\mathcal{O}}(k)^{\perp}\widehat{\mathcal{O}}(k) = 0$
7. Define
$$\widehat{\mathcal{T}}(k) := [\widehat{\mathcal{O}}(k)^{\perp}]^{\dagger}\widehat{\mathcal{O}}(k)^{\perp}H(y(k))[H(u(k))]^{\dagger}$$
$$\widehat{\mathcal{O}}_L(k) := [\text{ last } m(h_1-1) \text{ row blocks of } \widehat{\mathcal{O}}(k)]$$
$$\widehat{\mathcal{O}}_F(k+1) := [\text{ first } m(h_1-1) \text{ row blocks of } \widehat{\mathcal{O}}(k+1)]$$
$$\widehat{\mathcal{T}}_F(k) := [\text{ first } m \text{ columns } of\ \widehat{\mathcal{T}}(k)]$$
$$\widehat{\mathcal{T}}_L(k) := [\text{ last } p(h_1-1) \text{ row blocks } of\ \widehat{\mathcal{T}}_F(k)]$$
8. Define:
   - $A(k) := [\widehat{\mathcal{O}}_F(k+1)]^{\dagger}[\widehat{\mathcal{O}}_L(k)]$
   - $B(k) := [\mathcal{O}_L(k)]^{\dagger}\widehat{\mathcal{T}}_L(k)$
   - $C(k) :=$ [first p rows of $\widehat{\mathcal{O}}(k)$]
   - $D(k) :=$ [first p-by-m block of $\widehat{\mathcal{T}}(k)$]
9. Set $k = k + 1$, and start from Step 2

	Assumptions	Input/Output	Methodology
Parameter-based technique	•Slowly time-varying •Single response available	**Input:** Input-output sequences **Output:** Constant coefficients of an input-output model	Least squares
Basis function-based technique	•Suitable basis functions •Number of basis functions to be used •Suitable weights for WLS •Single response available	**Input:** Input-output sequences, basis function **Output:** Constant parameters of basis functions	Weighted least squares
Matrix case	•Access to impulse response data •Single response available	**Input:** Impulse responses **Output:** System matrices	Rank-revealing factorization
Subspace approach	•Slowly time-varying •Multiple responses available	**Input:** Hankel matrices of input-output data **Output:** System matrices	Singular value decomposition, QR decomposition

Table 2.1: Summary of LTV system identification techniques.

i.e. an ensemble of input responses is given by

$$\begin{bmatrix} y_j(k) & y_j(k+1) & \cdots & y_j(k+N-1) \\ y_{j+1}(k) & y_{j+1}(k+1) & \cdots & y_{j+1}(k+N-1) \\ \vdots & \vdots & \cdots & \vdots \\ y_{j+J-1}(k) & y_{j+J-1}(k+1) & \cdots & y_{j+J-1}(k+N-1) \end{bmatrix}$$

where k and N indicates the first and last time instance, respectively. j and J indicates the first and last experiment.

2.4 Summary and discussion

Table 2.1 summarises the linear time-varying system identification techniques discussed in this chapter.

A common assumption for the LTV system identification using LS or subspace methods is that the system is slowly time-varying. This assumption plays an important role in

their identification procedure, e.g. they are based in the idea that the time-varying system can be identified in time-invariant windows, allowing the procedure to track the time-varying parameters. When the system is fast time-varying, the use of basis functions is more suitable for accurate time-varying parameter identification, however, one needs to choose suitable basis function based on prior information of the system to be identified.

The matrix case in the other hand allows the extraction of system matrices of a state space representation for the data. However, the main assumption is that the impulse response is available, weather or not this is realizable. The availability of this information depends on the application.

In this thesis we present a duality-based approach for LTV system identification. Our approach is germane to the *subspace identification* approach. However, our technique differs essentially in that it does not exploit the structure and factorizations of Hankel matrices computed from the data. We instead exploit the fact that *external properties*- i.e. properties at the level of external variables, in our case *duality*- are reflected into *internal ones*- i.e. at the level of state. This allow us to develop algorithms to identify systems of fast and slow time-varying processes, as well as time-invariant ones.

Chapter 3

Linear time-varying systems

In this chapter, we introduce basic mathematical concepts of linear time-varying systems in the state space approach that we will use in the following chapters. This material is taken from different sources which are referenced through the chapter.

3.1 LTV state space representation

The class of systems considered in this thesis are described by a set of first order differential equations with time-varying coefficients of the form

$$\Sigma_{ss} \qquad \begin{aligned} \frac{d}{dt}x(\cdot) &= A(\cdot)x(\cdot) + B(\cdot)u(\cdot) \\ y(\cdot) &= C(\cdot)x(\cdot) + D(\cdot)u(\cdot) \end{aligned} \tag{3.1}$$

where $x(\cdot)$ is the *state* function of dimension n, $u(\cdot)$ is the *input* function of dimension m, and $y(\cdot)$ is the *output* function of dimension p. The matrices $A(\cdot) \in \mathcal{A}^{n\times n}$, $B(\cdot) \in \mathcal{A}^{n\times m}$, $C(\cdot) \in \mathcal{A}^{p\times n}$ and $D(\cdot) \in \mathcal{A}^{p\times m}$, whose value at $t \in \mathbb{R}$ is respectively $A(t)$, $B(t)$, $C(t)$ and $D(t)$, are called the *system matrices* of Σ_{ss}. To avoid confusion in notation, we consider $f(\cdot)$ to be a continuous function and $f(t)$ to be the value of the function $f(\cdot)$ at time t.

Consider the matrices $A(\cdot), B(\cdot), C(\cdot)$ and $D(\cdot)$ satisfying (3.1). Let $P(t) \in \mathbb{R}^{n\times n}$ be a nonsingular matrix for all $t \in \mathbb{R}$ satisfying

$$\frac{d}{dt}P(t) = A(t)P(t), \tag{3.2}$$

then a *transition matrix* of Σ_{ss} is defined as

$$\phi(t,\tau) := P(t)P(\tau)^{-1}, \tag{3.3}$$

where $P(t)$ is called *fundamental matrix*.

Some properties of the transition matrix are:

1. $\frac{d}{dt}\phi(t,\tau) = A(t)\phi(t,\tau)$
2. $\phi(t,t) = I$
3. $\phi(t,\tau)^{-1} = \phi(\tau,t)$
4. $\phi(t_1,\tau)\phi(\tau,t_2) = \phi(t_1,t_2)$.

These properties are a direct consequence of (3.2) and (3.3).

Let $x(t_0)$ for some arbitrary time t_0 be given, then the solution of Σ_{ss} for some $t > t_0$ is given by

$$x(t) = \phi(t,t_0)x(t_0) + \int_{t_0}^{t} \phi(t,\tau)B(\tau)u(\tau)d\tau, \tag{3.4}$$

see Kailath (1980) pp 601 eq. (21). It was also shown that the solution (3.4) can be easily verified by differentiating (3.4) with respect to t, i.e.

$$\frac{d}{dt}x(t) = \frac{d}{dt}\left(\phi(t,t_0)x(t_0)\right) + \frac{d}{dt}\left(\int_{t_0}^{t} \phi(t,\tau)B(\tau)u(\tau)d\tau\right),$$

and using (3.2) and (3.3),

$$\begin{aligned}
\frac{d}{dt}x(t) =& \left(\frac{d}{dt}P(t)\right)P(t_0)^{-1}x(t_0) + \left(\frac{d}{dt}P(t)\right)\int_{t_0}^{t} P(\tau)^{-1}B(\tau)u(\tau)d\tau \\
&+ P(t)\frac{d}{dt}\left(\int_{t_0}^{t} P(\tau)^{-1}B(\tau)u(\tau)d\tau\right) \\
=& A(t)P(t)P(t_0)^{-1}x(t_0) + A(t)P(t)\int_{t_0}^{t} P(\tau)^{-1}B(\tau)u(\tau)d\tau \\
&+ P(t)P(t)^{-1}B(t)u(t) \\
=& A(t)\left(P(t)P(t_0)^{-1}x(t_0) + P(t)\int_{t_0}^{t} P(\tau)^{-1}B(\tau)u(\tau)d\tau\right) + B(t)u(t) \\
=& A(t)\left(\phi(t,t_0)x(t_0) + \int_{t_0}^{t} \phi(t,\tau)B(\tau)u(\tau)d\tau\right) + B(t)u(t) \\
=& A(t)x(t) + B(t)u(t).
\end{aligned}$$

The solution of the output equation of Σ_{ss} at time t is then given by

$$y(t) = C(t)\left(\phi(t,t_0)x(t_0) + \int_{t_0}^{t} \phi(t,\tau)B(\tau)u(\tau)d\tau\right) + D(t)u(t). \tag{3.5}$$

3.1.1 Equivalent transformations

In identification theory, it is often important to consider equivalent transformation of a system for analysis. This transformation is basically a change of state coordinates in which the input-output relation is not affected.

Definition 3.1. (Silverman (1966) p. 15) Let $T(t) \in \mathbb{R}^{n\times n}$ be a time-varying matrix, nonsingular for all $t \in \mathbb{R}$, and let $A(\cdot)$, $B(\cdot)$, $C(\cdot)$ and $D(\cdot)$ be the system matrices of (3.1). Define $z(\cdot) := T(\cdot)x(\cdot)$ and matrices

$$\begin{aligned}\overline{A}(\cdot) &:= T(\cdot)A(\cdot)T(\cdot)^{-1} - T(\cdot)\frac{d}{dt}T(\cdot)^{-1}\\ \overline{B}(\cdot) &:= T(\cdot)B(\cdot)\\ \overline{C}(\cdot) &:= C(\cdot)T(\cdot)^{-1}\\ \overline{D}(\cdot) &:= \overline{D}(\cdot)\end{aligned} \tag{3.6}$$

satisfying

$$\begin{aligned}\frac{d}{dt}z(\cdot) &= \overline{A}(\cdot)z(\cdot) + \overline{B}(\cdot)u(\cdot)\\ y(\cdot) &= \overline{C}(\cdot)z(\cdot) + \overline{D}(\cdot)u(\cdot).\end{aligned} \tag{3.7}$$

Then the system (3.7) is said to be *equivalent* to the system (3.1).

Def. 3.1 is illustrated in the following. Let the state variable transformation of Def. 3.1 be given, i.e.

$$z(\cdot) := T(\cdot)x(\cdot). \tag{3.8}$$

Now we apply the variable transformation to the system (3.1). From (3.8) we have that $x(\cdot) = T(\cdot)^{-1}z(\cdot)$. Substituting $x(\cdot)$ with $T(\cdot)^{-1}z(\cdot)$ in (3.1) we have

$$\begin{aligned}\frac{d}{dt}(T(\cdot)^{-1}z(\cdot)) &= A(\cdot)(T(\cdot)^{-1}z(\cdot)) + B(\cdot)u(\cdot)\\ y(\cdot) &= C(\cdot)(T(\cdot)^{-1}z(\cdot)) + D(\cdot)u(\cdot).\end{aligned}$$

Differentiating the left hand side of the above equation

$$\begin{aligned}\left(\frac{d}{dt}T(\cdot)^{-1}\right)z(\cdot) + T(\cdot)^{-1}\frac{d}{dt}z(\cdot) &= A(\cdot)(T(\cdot)^{-1}z(\cdot)) + B(\cdot)u(\cdot)\\ y(\cdot) &= C(\cdot)(T(\cdot)^{-1}z(\cdot)) + D(\cdot)u(\cdot).\end{aligned}$$

From such equality, the transformed system with the new state variable $z(\cdot)$ is represented by

$$\begin{aligned}\frac{d}{dt}z(\cdot) &= \left(T(\cdot)A(\cdot)T(\cdot)^{-1} - T(\cdot)\frac{d}{dt}T(\cdot)^{-1}\right)z(\cdot) + (T(\cdot)B(\cdot))u(\cdot)\\ y(\cdot) &= \left(C(\cdot)T(\cdot)^{-1}\right)z(\cdot) + D(\cdot)u(\cdot).\end{aligned} \tag{3.9}$$

Note that the output generated by applying an input to Σ_{ss} is the same as the output generated by applying the same input to (3.7). However, the state trajectories between these systems are not the same.

3.1.2 Controllability and observability

In the literature, controllability and observability for LTV systems can be defined partially or totally, i.e. a system may not be controllable (observable) in some interval of time, but it can be controllable (observable) in some subinterval, see Silverman (1966). However, in the case when one considers analytic functions of time, the distinction between partial controllability and total controllability is abandoned. The reason is that if a matrix with analytic entries has rank n at a finite number of points on an interval, then it has rank n in every point on that interval, see Chen (1967).

In the following we introduce controllability and observability based on Silverman (1966). The functions are considered to be analytic.

Definition 3.2. (Silverman (1966)) The system Σ_{ss} is *controllable*, if for any state x_0 at t_0, and any desired final state x_1 at t_1, there exist an input $u(\cdot)$ and a trajectory $x(\cdot)$ such that $x(t_1) = x_1$.

Now consider the impulse response of Σ_{ss} as follows. Define

$$H(t,\tau) := \begin{cases} C(t)\phi(t,\tau)B(\tau), & t > \tau \\ C(t)\phi(t,\tau)B(\tau) + D(t), & t = \tau \\ 0, & t < \tau \end{cases} \tag{3.10}$$

to be the impulse response matrix of Σ_{ss}. From (3.3) and (3.10), we can define

$$\Psi(t) := C(t)P(t) \tag{3.11}$$

$$\Theta(\tau) := P(\tau)^{-1}B(\tau), \tag{3.12}$$

such that the impulse response matrix can be written as

$$H(t,\tau) = \Psi(t)\Theta(\tau) \quad \text{for } t > \tau, \tag{3.13}$$

with $P(t)$ being a fundamental matrix of Σ_{ss}.

It can be shown that the impulse response is invariant under equivalent transformation. The reason is that if $P(t)$ is a fundamental matrix of Σ_{ss}, then $T(t)P(t)$ is a fundamental matrix of $\overline{\Sigma}_{ss}$ such that $\overline{\phi}(t,\tau) = T(t)\phi(t,\tau)T(\tau)^{-1}$, see p. 16 of Silverman (1966). Let $\overline{P}(t) := T(t)P(t)$ be a fundamental matrix of $\overline{\Sigma}_{ss}$, it follows that the impulse response

for $\overline{\Sigma}_{ss}$ can be written as

$$\overline{H}(t,\tau) = \overline{C}(t)\overline{\phi}(t,\tau)\overline{B}(\tau) \quad t > \tau,$$

where $\overline{\phi}(t,\tau) = T(t)\phi(t,\tau)T(\tau)^{-1}$, then

$$\begin{aligned}\overline{H}(t,\tau) &= C(t)T(t)^{-1}T(t)\phi(t,\tau)T(\tau)^{-1}T(\tau)B(\tau)\\ &= C(t)\phi(t,\tau)B(\tau)\\ &= H(t,\tau).\end{aligned}$$

Let $P(\cdot)$ be a fundamental matrix for Σ_{ss}, and consider $\Theta(\cdot) = P(\cdot)^{-1}B(\cdot)$. The following theorem is a result of Silverman (1966).

Theorem 3.3. *(Silverman (1966)) The system Σ_{ss} is controllable if and only if the rows of $\Theta(\cdot)$ are linearly independent functions*[1]*.*

It is important to note that controllability is valid for any equivalent transformation as follows. Let $\overline{P}$ be a fundamental matrix for the system (3.7), i.e. $\overline{P}(t) = T(t)P(t)$. Let

$$\overline{\Theta}(t) := \overline{P}(t)^{-1}\overline{B}(t). \tag{3.14}$$

From Def. 3.1, $\overline{B}(t) = T(t)B(t)$. It follows that

$$\begin{aligned}\overline{P}^{-1}\overline{B}(t) &= P(t)^{-1}T(t)^{-1}T(t)B(t)\\ &= P(t)^{-1}B(t)\end{aligned}$$

which implies that $\overline{\Theta}(t) = \Theta(t)$.

The ability to determine the state at particular time from the system outputs is called *observability*. Analogous to controllability of LTV systems, we now introduce observability of LTV systems.

Definition 3.4. (Silverman (1966)) The system Σ_{ss} is *observable*, if any initial state x_0 at t_0 can be determined from the system output and input.

Let $P(\cdot)$ be a fundamental matrix of Σ_{ss}, and $\Psi(\cdot) = C(\cdot)P(\cdot)$. Then observability conditions are given by the following theorem, see Silverman and Meadows (1967) p. 65.

Theorem 3.5. *(Silverman and Meadows (1967)) The system Σ_{ss} is observable if and only if the columns of $\Psi(\cdot)$ are linearly independent.*

[1] See Definition A.1 in Appendix A.

3.1.2.1 Controllability matrix

In the following, a characterization of the controllability matrix is given in terms of the system matrices based on the results in Silverman and Meadows (1967). Define

$$P_0(\cdot) := B(\cdot) \quad P_{s+1} := -A(\cdot)P_s(\cdot) + \frac{d}{dt}P_s(\cdot) \quad s = 0, .., \ell - 2,$$

then the controllability matrix for Σ_{ss} is defined by

$$\mathrm{C}_\ell(\cdot) := \begin{bmatrix} P_0(\cdot) & P_1(\cdot) & \cdots & P_{\ell-1}(\cdot) \end{bmatrix}. \tag{3.15}$$

Initially it may not be evident that C_ℓ is related to controllability in the sense of Th. 3.3. To see how this is related to controllability we need first to consider the *Wronskian matrix.*

Definition 3.6. (Silverman and Meadows (1967)) Let $F(\cdot)$ be an n-dimensional column vector of functions which are (n-1) times differentiable, i.e.

$$F(\cdot) := \begin{bmatrix} f_1(\cdot) \\ f_2(\cdot) \\ \vdots \\ f_n(\cdot) \end{bmatrix},$$

then the Wronskian matrix for $F(\cdot)$ is defined as

$$W(\cdot) := \begin{bmatrix} F(\cdot) & \frac{d}{dt}F(\cdot) & \cdots & \frac{d^{n-1}}{dt^{n-1}}F(\cdot) \end{bmatrix}.$$

Proposition 3.7. *(Silverman and Meadows (1967)) Let $F(\cdot)$ be defined as in Def. 3.6. If the Wronskian matrix of $F(\cdot)$ has rank n as a set of functions for some t, then the rows of $F(\cdot)$ are linearly independent as a set of functions.*

Proof. See Lemma 1 p. 67 of Silverman and Meadows (1967). □

Now consider the Wronskian matrix of (3.12) of system Σ_{ss},

$$W_\Theta(\cdot) := \begin{bmatrix} \Theta & \frac{d}{dt}\Theta & \cdots & \frac{d^{n-1}}{dt^{n-1}}\Theta \end{bmatrix}. \tag{3.16}$$

If $T(\cdot)$ is a fundamental matrix of Σ_{ss}, it follows that $W_\Theta(\cdot)$ can be written as

$$W_\Theta(\cdot) = T(\cdot)^{-1}\mathrm{C}_n(\cdot),$$

see Property 3.1 in p. 26 of Silverman (1966).

Since $T(\cdot)$ is nonsingular for all $t \in \mathbb{R}$, then the rank of $W_\Theta(\cdot)$ is equal to the rank of $\mathrm{C}_n(\cdot)$. Then the following theorem holds.

Theorem 3.8. *(Silverman and Meadows (1967) p. 70) The system Σ_{ss} is controllable, iff $\mathrm{C}_n(\cdot)$ has rank n.*

3.2 From numbers to functions

The data used for system identification is gathered as a sequence of samples taken at a certain time. Therefore, to develop a system identification algorithm we need to make the connection between sequences of data and continuous functions of time. In this section we use *Reproducing kernel Hilbert spaces* (RKHS) to address this problem. We will introduce the theory and notation that we will use to embed RKHS into our identification approach.

Basic knowledge of functional analysis is required for this Chapter which is provided in appendix A. The theory of RKHS presented in this chapter can be found in Steinwart and Christmann (2008).

RKHS has shown itself to be a powerful tool for learning algorithms, e.g. finding predictive functions based on data, see Pillonetto et al. (2014). The advantages of using RKHS is that we can predefine properties of the function we are going to deal with, e.g. infinitely differentiability, by choosing appropriate kernels.

3.2.1 Kernels

In this section, we introduce the notion of *kernels*.

Definition 3.9. (Steinwart and Christmann (2008) p. 112) Let X be a non-empty set. Then a symmetric *form*[2] $k : X \times X \to \mathbb{R}$ is called a kernel on X if there exist a *Hilbert space*[3] H and a map $\Phi : X \to H$ s.t. for all $x, x' \in X$

$$k(x, x') = \langle \Phi(x), \Phi(x') \rangle_H \tag{3.17}$$

holds.

Φ is called a feature map and H a feature space of k. Moreover, note that because of symmetry $k(x, x') = k(x', x)$.

Given a kernel, the feature map and feature space associated to it is not uniquely determined. This is illustrated in the following trivial example, see Silverman and Meadows (1967).

[2] See Definition A.19 in Appendix A.
[3] See Definition A.14 in Appendix A.

Example 3.1. *Let $X := \mathbb{R}$ and define a linear kernel as $k(x, x') = xx'$ for all $x, x' \in X$. Define the map $\Phi : X \to \mathbb{R}^2$ as $\Phi(x) := [\frac{x}{\sqrt{2}}, \frac{x}{\sqrt{2}}]^\top$ for all $x \in X$. Then Φ is a feature map of k since*

$$\langle \Phi(x), \Phi(x') \rangle_H = \frac{x}{\sqrt{2}} \frac{x'}{\sqrt{2}} + \frac{x}{\sqrt{2}} \frac{x'}{\sqrt{2}} = xx'. \tag{3.18}$$

with feature space $\mathbb{R}^2$.

From example 3.1 it is evident that feature maps and feature spaces are not unique, however, kernels are unique. For example, we can define $\Phi(x) := x$ in example 3.1, and the resulting kernel will coincide with that in (3.18). Note also that, in this case, the feature space corresponding to $\Phi(x) = x$ is now $\mathbb{R}$.

Remark 3.10. In this work we assume that kernels are given. The reader is referred to Aronszajn (1950) for further details about the construction of kernels.

In order to determine whether or not a function is a kernel, we need to find a corresponding feature space for the kernel. However, this can be difficult in some cases, e.g. when the feature space is infinite dimensional. Another alternative is to characterize kernels as follows, see Steinwart and Christmann (2008) page 117.

Definition 3.11. (Steinwart and Christmann (2008) p. 177) A function $k : X \times X \to \mathbb{R}$ is positive definite if for all $n \in \mathbb{N}$ and all $x_1, .., x_n \in X$, $\alpha_1, .., \alpha_n \in \mathbb{R}$, the following holds

$$\sum_{i=1}^{n} \sum_{j=1}^{n} \alpha_i \alpha_j k(x_i, x_j) \geq 0. \tag{3.19}$$

The function k is strictly positive definite if, for mutually distinct $x_1, .., x_n \in X$, (3.19) only holds for $\alpha_1 = \cdots = \alpha_n = 0$.

Remark 3.12. In the literature of learning theory, researchers use the term "positive definite" to refer to "positive semi-definiteness", and "strictly positive definite" to refer to "positive definiteness", see note after definition 4.15 in Steinwart and Christmann (2008) page 117.

As we can see in example 3.1, kernels are inner products between feature maps. Since inner products are positive definite, then the kernel must also be positive definite. If a kernel is not positive definite, it may not represent an inner product in a Hilbert space. The following theorem states sufficient conditions for a functional k to be a kernel without explicitly define feature maps.

Theorem 3.13. *(Steinwart and Christmann (2008)) A functional $k : X \times X \to \mathbb{R}$ is a kernel if it is symmetric and positive definite.*

Proof. See Theorem 4.16 of Steinwart and Christmann (2008) in page 118. □

We can also define new kernels using linear combination of given kernels. This is stated in the following proposition.

Proposition 3.14. *(Shawe-Taylor and Cristianini (2004)) Let X be a normed vector space*[4]*. Let k_1 and k_2 be kernels over $X \times X$, $a \in \mathbb{R}$. Then for $x, y \in X$ the following functions are kernels:*

1. $k(x, y) := k_1(x, y) + k_2(x, y)$,
2. $k(x, y) := ak_1(x, y)$,
3. $k(x, y) := k_1(x, y)k_2(x, y)$,
4. $k(x, y) := \exp(k_1(x, y))$.

Proof. See Shawe-Taylor and Cristianini (2004) pages 75-76. □

Remark 3.15. In the actual literature, there exist a variety of kernels that have been developed for different applications depending on the field of study. Some examples of such kernels include: the linear kernel, polynomial kernel and Gaussian kernel. The kernel can be chosen depending on the application and the nature of the data to be described.

3.2.2 Reproducing kernel Hilbert spaces

Now we introduce RKHS and show how a reproducing kernel defines a unique RKHS associated to it based on Steinwart and Christmann (2008).

Definition 3.16. Let $\mathcal{H} \subset \mathbb{R}^X$ be a Hilbert space (not necessarily RKHS) defined on a non-empty set X. A form $k : X \times X \to \mathbb{R}$ is called a ***reproducing kernel*** of $\mathcal{H}$ if for all $x, y \in X$ and all $f \in \mathcal{H}$ the following conditions are satisfied:

1. $k(\cdot, x) \in \mathcal{H}$
2. $\langle f, k(\cdot, x)\rangle_{\mathcal{H}} = f(x)$, (*reproducing property*),
3. $k(x, y) = \langle k(\cdot, x), k(\cdot, y)\rangle_{\mathcal{H}}$
4. $k(x, y) = k(y, x)$ (*symmetry*).

Moreover, the space $\mathcal{H}$ is called a ***reproducing kernel Hilbert space*** over X if for all $x \in X$ there exists a unique $\mathcal{F}_x : \mathcal{H} \longrightarrow \mathbb{R}$ such that $\mathcal{F}_x(f) = f(x)$, and it is *continuous*[5].

[4] See Definition A.9 in Appendix A.
[5] See Definition A.22 in Appendix A.

Note that $k(\cdot, x)$ maps $x \mapsto g$ where $x \in X$ and $g \in \mathcal{H}$, and that $\mathcal{F}_x(f) = \langle f, k(\cdot, x)\rangle_{\mathcal{H}} = f(x)$, the reproducing property. Note also that the adjective *reproducing* is added to indicate that the kernel has the reproducing property defined in condition 2 of Def. 3.16.

Consider a given finite set of elements in X, i.e. $x_i \in X$ for $i = 1, .., l$. Given a reproducing kernel k, one can construct the RKHS associated to it as the space of functions spanned by $\{k(\cdot, x)|x \in X\}$. Then every function in this space is represented as a linear combination of $k(\cdot, x)$. This is stated in Schölkopf et al. (2001) as the *Representer theorem.*

Theorem 3.17. *(Schölkopf et al. (2001) Representer theorem) Every function in the RKHS can be represented as a linear combination of the kernel of $\mathcal{H}$, i.e. $\exists \alpha_i \in \mathbb{R}$ such that*

$$f(\cdot) = \sum_{i=1}^{l} \alpha_i k(\cdot, x_i), \tag{3.20}$$

where $i, l \in \mathbb{N}$, $x_1, .., x_l \in X$, and $\alpha_1, .., \alpha_l \in \mathbb{R}$.

Proof. See Th. 4 of Schölkopf et al. (2001) page 5. □

Then the *Moore-Aronszajn theorem* states the following important result.

Theorem 3.18. *(Aronszajn (1950)) Let $k(\cdot, x)$ be a given positive definite reproducing kernel. Let f and g be functions defined as*

$$f(\cdot) := \sum_{i=1}^{l} \alpha_i k(\cdot, x_i),$$
$$g(\cdot) := \sum_{j=1}^{l} \beta_j k(\cdot, x_j).$$

Then k defines a unique RKHS

$$\mathcal{H} := \{f(\cdot) \mid f(\cdot) = \sum_{i=1}^{l} \alpha_i k(\cdot, x_i)\} \tag{3.21}$$

equipped with the inner product

$$\langle f, g\rangle_{\mathcal{H}} = \sum_{i=1}^{l} \sum_{j=1}^{l} \alpha_i \beta_j k(x_i, x_j). \tag{3.22}$$

It follows from Th. 3.18 that given a reproducing kernel we can define a unique RKHS associated to it. Then from the reproducing property in Def. (3.16) we have that

$$\langle f, k(\cdot, x_j)\rangle_{\mathcal{H}} = \sum_{i=1}^{l} \alpha_i k(x_j, x_i) = f(x_j)$$

for given $x_1, ., , x_l \in X$.

We can see that by the reproducing property, we can evaluate the function f at a point x_j by applying the inner product between f and k.

3.2.2.1 Computation of $\alpha's$

In this section we will show how the coefficients of the functions in a RKHS are computed.

Let $f(x_i) \in Y$ and $x_i \in X$ for $i = 1, .., N$ be given. It follows from theorem 3.18 that for a given kernel k the RKHS is given by

$$\mathcal{H} := \{f(\cdot) \mid f(\cdot) = \sum_{i=1}^{N} \alpha_i k(\cdot, x_i)\}.$$

It follows that $f(\cdot) \in \mathcal{H}$ evaluated at $x' \in X$ is

$$f(x') = \sum_{i=1}^{N} \alpha_i k(x', x_i). \tag{3.23}$$

Since $f(x_i)$ are given, then the constant coefficients $\alpha's$ are obtained by solving the following linear equation

$$f(x_j) = \sum_{i=1}^{N} \alpha_i k(x_j, x_i) \quad j = 1, .., N. \tag{3.24}$$

We can write (3.24) in matrix form as

$$\begin{bmatrix} f(x_1) \\ f(x_2) \\ \vdots \\ f(x_N) \end{bmatrix} = \begin{bmatrix} k(x_1, x_1) & k(x_1, x_2) & \cdots & k(x_1, x_N) \\ k(x_2, x_1) & k(x_2, x_2) & \cdots & k(x_2, x_N) \\ \vdots & \vdots & \ddots & \vdots \\ k(x_N, x_1) & k(x_N, x_2) & \cdots & k(x_N, x_N) \end{bmatrix} \begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_N \end{bmatrix}. \tag{3.25}$$

Defining

$$\begin{aligned} \mathbf{Y}_{x_1,x_N} &:= \text{col}(f(x_1), .., f(x_N)) \\ \mathbf{K}_{x_1,x_N} &:= [k(x_i, x_j)]_{i,j} \quad i, j = 1, .., N \\ \mathbf{A}_{1,N} &:= \text{col}(\alpha_1, .., \alpha_N) \end{aligned}$$

we can write (3.25) compactly as

$$\mathbf{Y}_{x_1,x_N} = \mathbf{K}_{x_1,x_N} \mathbf{A}_{1,N} \tag{3.26}$$

where $\mathbf{Y}_{x_1,x_N} \in \mathbb{R}^N$, $\mathbf{K}_{x_1,x_N} \in \mathbb{R}^{N\times N}$ and $\mathbf{A}_{1,N} \in \mathbb{R}^N$. Then we can solve eq. (3.26) for $\mathbf{A}_{1,N}$.

Note that $\mathbf{K}_{x_1,x_N}$ in (3.26) is symmetric. The solution of eq. (3.26) is guaranteed by Def. 3.9: since the kernel is positive definite then the matrix $\mathbf{K}_{x_1|x_N}$ associated to the kernel is positive definite and consequently nonsingular.

We now introduce the differentiability of the functions in a RKHS. Consider the following result taken from Steinwart and Christmann (2008).

Theorem 3.19. *(Steinwart and Christmann (2008)) Let k be an m-times continuously differentiable kernel with its associated RKHS $\mathcal{H}$. Then every $f \in \mathcal{H}$ is m-times continuously differentiable.*

Proof. see proof of Corollary 4.36 p. 131 of Steinwart and Christmann (2008). □

3.3 Summary

In this chapter we discussed some properties of linear time-varying systems in the state space approach. We covered equivalent transformations. We also defined controllability and observability for state space LTV representations. The theory of reproducing kernel Hilbert spaces is also introduced. We outlined the relationship between kernels and RKHS that are needed for the construction of RKHS.

Chapter 4

External structures and internal representations

In this chapter duality is introduced. We show how external properties at the level of inputs and outputs are reflected in the internal ones at the level of the state. We also define special properties as self-adjointness that can be presented in certain cases. We discuss the connection between them and their influence in duality.

4.1 The adjoint system

The study of duality of linear time-invariant and time-varying systems has a long history, see e.g. Kalman (1960). For the time-varying case, a module-theoretic approach to duality is developed in Rudolph (1996). We introduce the following intrinsic trajectory-based definition of adjoint system, inspired by Crouch et al. (1995); Van der Schaft (1991).

Definition 4.1. (Crouch et al. (1995); Van der Schaft (1991))Let Σ, Σ' be two LTV systems with input-state-output variables (u, x, y) and (u', x', y'), respectively. Σ and Σ' are called *adjoint* if there exists $Q(\cdot) = Q(\cdot)^\top$ nonsingular everywhere on $\mathbb{R}$, such that for all trajectories $(x(\cdot), u(\cdot), y(\cdot))$ of Σ and $(x'(\cdot), u'(\cdot), y'(\cdot))$ of Σ' it holds that

$$u^\top y' + y^\top u' = \frac{d}{dt}\left(x^\top Q x'\right) \ . \tag{4.1}$$

If $Q(\cdot) = I_n$ in (4.1), we call the i-s-o representations induced by $(A(\cdot), B(\cdot), C(\cdot), D(\cdot))$ and $(A'(\cdot), B'(\cdot), C'(\cdot), D'(\cdot))$ *matched*. Throughout the thesis we will equivalently use the term *adjoint* and *dual*.

We now introduce the integral of the external structure, i.e. inputs and outputs of the primal and its adjoint. In order to make sure that such integral exists, we assume that the trajectories have compact support.

Proposition 4.2. *(Van der Schaft (1991)) Let (x, u, y) and (x', u', y') be compact-support trajectories[1] of an i-s-o system Σ and its adjoint (dual) Σ', then it holds that*

$$\int_{-\infty}^{+\infty} u(\tau)^\top y'(\tau) + y(\tau)^\top u'(\tau)\ d\tau = 0. \tag{4.2}$$

Proof. Assume that (x, u, y) and (x', u', y') have support in the interval (a, b). From (4.1), it follows that

$$\int_a^b u(\tau)^\top y'(\tau) + y(\tau)^\top u'(\tau)\ d\tau = x(b)^\top Q(b)x'(b) - x(a)^\top Q(a)x'(a).$$

It follows that $x(b) = x'(b) = x(a) = x'(a) = 0$ and consequently (4.2) holds. This yields the claim. □

In order to characterize the relationship between the system matrices of the primal and its dual, we provide the following definition.

Definition 4.3. Let $(A(\cdot), B(\cdot), C(\cdot), D(\cdot))$ be an i-s-o representation of a LTV system. $(A(\cdot), B(\cdot), C(\cdot), D(\cdot))$ is *trim* if $\forall v \in \mathbb{R}^{n+m}$ there exists a trajectory $\mathrm{col}(\widetilde{x}(\cdot), \widetilde{u}(\cdot))$ and $t \in \mathbb{R}$ satisfying (3.1) such that $\mathrm{col}(\widetilde{x}(t), \widetilde{u}(t)) = v$.

The intuition behind trimness is that through every point in $\mathbb{R}^{n+m}$ there is a system trajectory passing through it.

Let us denote $\dot{M}(\cdot)$ as the matrix whose (i, j)-th entry is the first derivative of the (i, j)-th entry of the matrix $M(\cdot)$. The following result shows how the system matrices of the primal and dual system are related.

Proposition 4.4. *Let $(A(\cdot), B(\cdot), C(\cdot), D(\cdot))$ and $(A'(\cdot), B'(\cdot), C'(\cdot), D'(\cdot))$ be trim i-s-o representation of a system and its adjoint. Let Q in (4.1) be such that $Q(\cdot) = Q(\cdot)^\top$ nonsingular everywhere on $\mathbb{R}$, then*

$$\begin{aligned} Q(\cdot)A'(\cdot) + A(\cdot)^\top Q(\cdot) + \dot{Q}(\cdot) &= 0_{n\times n} \\ Q(\cdot)B'(\cdot) - C(\cdot)^\top &= 0_{n\times p} \\ B(\cdot)^\top Q(\cdot) - C'(\cdot) &= 0_{m\times n} \\ D'(\cdot) + D(\cdot)^\top &= 0_{m\times p}\,. \end{aligned} \tag{4.3}$$

[1] See Definition A.6 in Appendix A.

Proof. Write (4.1) using the i-s-o equations for Σ and Σ':

$$\begin{aligned} & u(\cdot)^\top \left(C'(\cdot)x'(\cdot) + D'(\cdot)u'(\cdot)\right) + \left(C(\cdot)x(\cdot) + D(\cdot)u(\cdot)\right)^\top u'(\cdot) \\ = \; & \left(A(\cdot)x(\cdot) + B(\cdot)u(\cdot)\right)^\top Q(\cdot)x'(\cdot) + x(\cdot)^\top \dot{Q}(\cdot)x'(\cdot) \\ & + x(\cdot)^\top Q(\cdot)\left(A'(\cdot)x'(\cdot) + B'(\cdot)u'(\cdot)\right) \; . \end{aligned}$$

Now, factorize $x(\cdot)^\top$, $u(\cdot)^\top$, $x'(\cdot)$ and $u'(\cdot)$ in the above equation.

$$\begin{aligned} & u(\cdot)^\top(-C'(\cdot) + B(\cdot)^\top Q(\cdot))x'(\cdot) + u(\cdot)^\top(-D'(\cdot) - D(\cdot)^\top)u'(\cdot) \\ & + x(\cdot)^\top(-C(\cdot)^\top + Q(\cdot)B'(\cdot))u'(\cdot) \\ & + x(\cdot)^\top(A(\cdot)^\top Q(\cdot) + \dot{Q}(\cdot) + Q(\cdot)A'(\cdot))x'(\cdot) = 0. \end{aligned}$$

It is easy to see that such equality is equivalent in matrix form to

$$\begin{bmatrix} x(\cdot) \\ u(\cdot) \end{bmatrix}^\top \begin{bmatrix} A(\cdot)^\top Q(\cdot) + Q(\cdot)A'(\cdot) + \dot{Q}(\cdot) & Q(\cdot)B'(\cdot) - C(\cdot)^\top \\ B(\cdot)^\top Q(\cdot) - C'(\cdot) & -D'(\cdot) - D(\cdot)^\top \end{bmatrix} \begin{bmatrix} x'(\cdot) \\ u'(\cdot) \end{bmatrix} = 0 \; .$$

Now, with Def. 4.3, use trimness of the two i-s-o representations to conclude that for every pair of vectors $v \in \mathbb{R}^{n+m}$ and $v' \in \mathbb{R}^{n+p}$ there exist $t \in \mathbb{R}$ and trajectories $\mathrm{col}(x, u)$ and $\mathrm{col}(x', u')$ whose value at t is v, respectively v'. Since for every point in $\mathbb{R}^{n+m}$ and $\mathbb{R}^{n+p}$ there are trajectories of the primal and dual, respectively, passing through it, then for the above expression to be true the central matrix must be identically zero. This concludes the proof. □

Now we introduce the following result that is a consequence of Prop. 4.4.

Corollary 4.5. *If the trim i-s-o representations of a system and its adjoint associated respectively with* $(A(\cdot), B(\cdot), C(\cdot), D(\cdot))$ *and* $(A'(\cdot), B'(\cdot), C'(\cdot), D'(\cdot))$ *are matched, then*

$$\begin{aligned} A'(\cdot) + A(\cdot)^\top &= 0_{n\times n} \\ B'(\cdot) - C(\cdot)^\top &= 0_{n\times p} \\ D'(\cdot) + D(\cdot)^\top &= 0_{m\times p} \; . \end{aligned} \tag{4.4}$$

Proof. The proof follows from the definition of matched representations and Prop. 4.4. □

It follows from Cor. 4.5 that a matched representation of the adjoint system is

$$\begin{aligned} \frac{d}{dt}x'(\cdot) &= -A(\cdot)^\top x'(\cdot) + C(\cdot)^\top u'(\cdot) \\ y'(\cdot) &= B(\cdot)^\top x'(\cdot) - D(\cdot)^\top u'(\cdot) \; . \end{aligned} \tag{4.5}$$

where $x' \in \mathbb{R}^n$, $u' \in \mathbb{R}^p$ and $y' \in \mathbb{R}^m$. Now the corresponding dual state space, dual input space and dual output space can be identified with $\mathbb{X}_a := \mathbb{R}^n$, $\mathbb{U}_a := \mathbb{R}^p$ and $\mathbb{Y}_a := \mathbb{R}^m$ respectively. Note that the dimension of the dual input space is p and the dual output space is m; in contrast to the primal system, where the dimension of the input space and output space is m and p respectively.

We now consider state transformations in duality in the following result of this section.

Proposition 4.6. *Let $T(\cdot), T'(\cdot)$ be $n \times n$ matrices with analytic entries and everywhere nonsingular. Let $\mathrm{col}(u, y, x)$ and $\mathrm{col}(u', y', x')$ satisfy matched i-s-o equations of a system Σ and its dual Σ'. Define $z(\cdot) := T(\cdot)x(\cdot)$, $z'(\cdot) := T'(\cdot)x'(\cdot)$, and*

$$\widehat{Q}(\cdot) := T(\cdot)^{-\top} T'(\cdot)^{-1}. \tag{4.6}$$

Then

$$u^\top y' + y^\top u' = \frac{d}{dt}\left(z^\top \widehat{Q} z'\right) . \tag{4.7}$$

Proof. Since the system is matched we write (4.1) using $Q(\cdot) := I_n$:

$$u^\top y' + y^\top u' = \frac{d}{dt}(x^\top x').$$

Now using equalities $x(\cdot) = T(\cdot)^{-1} z(\cdot)$ and $x'(\cdot) = T'(\cdot)^{-1} z'(\cdot)$ it follows that

$$u^\top y' + y^\top u' = \frac{d}{dt}(z^\top T^{-\top} T'^{-1} z').$$

Now define $\widehat{Q}(\cdot) := T(\cdot)^{-\top} T'(\cdot)^{-1}$. This concludes the proof. □

From Prop. 4.6, if we consider a change of basis in a matched representation, e.g. $z(\cdot) = T(\cdot)x(\cdot)$ and $z'(\cdot) = T'(\cdot)x'(\cdot)$, then the corresponding representations with the new state variables are not matched. This means that there always exists an i-s-o representation in another basis in which the system is not matched. Now let us consider the special case when $T'(\cdot) = T(\cdot)^{-\top}$ in eq. (4.6). Then the following result holds.

Proposition 4.7. *Let $T(\cdot), T'(\cdot)$ be $n \times n$ matrices with analytic entries and everywhere nonsingular. Let $(A(\cdot), B(\cdot), C(\cdot), D(\cdot))$ and $(-A(\cdot)^\top, C(\cdot)^\top, B(\cdot)^\top, -D(\cdot)^\top)$, respectively, induce matched i-s-o representations of a system Σ and its adjoint Σ'.*

Define $z(\cdot) := T(\cdot)x(\cdot)$ and $z'(\cdot) := T'(\cdot)x'(\cdot)$. Then the i-s-o equations in the new state variables z, z' are associated with the matrices

$$\begin{aligned} A_z(\cdot) &:= T(\cdot)A(\cdot)T(\cdot)^{-1} + \dot{T}(\cdot)T(\cdot)^{-1} \\ B_z(\cdot) &:= T(\cdot)B(\cdot) \\ C_z(\cdot) &:= C(\cdot)T(\cdot)^{-1} \\ D_z(\cdot) &:= D(\cdot) \end{aligned}$$

and

$$\begin{aligned} A'_{z'}(\cdot) &:= -T'(\cdot)A(\cdot)^\top T'(\cdot)^{-1} + \overset{\bullet}{T'}(\cdot)T'(\cdot)^{-1} \\ B'_{z'}(\cdot) &:= T'(\cdot)C(\cdot)^\top \\ C'_{z'}(\cdot) &:= B(\cdot)^\top T'(\cdot)^{-1} \\ D'_{z'}(\cdot) &:= -D(\cdot)^\top , \end{aligned}$$

respectively. If $T'(\cdot) = T(\cdot)^{-\top}$, then the new i-s-o representations are also matched.

Define $\overset{\bullet}{\overline{M(\cdot)N(\cdot)}} := \overset{\bullet}{M}(\cdot)N(\cdot) + M(\cdot)\overset{\bullet}{N}(\cdot)$ and denote $\overset{\bullet}{\overline{M(\cdot)^{-1}}}$ as the matrix whose (i,j)-th entry is the first derivative of the (i,j)-th entry of $M(\cdot)^{-1}$.

Proof. Applying a variable transformation $x(\cdot) = T(\cdot)^{-1}z(\cdot)$ to the i-s-o representation of the system Σ, it follows that

$$\begin{aligned} \frac{d}{dt}(T(\cdot)^{-1}z(\cdot)) &= A(\cdot)(T(\cdot)^{-1}z(\cdot)) + B(\cdot)u(\cdot) \\ y(\cdot) &= C(\cdot)(T(\cdot)^{-1}z(\cdot)) + D(\cdot)u(\cdot). \end{aligned}$$

Let $\overset{\bullet}{\overline{T(\cdot)^{-1}}} := \frac{d}{dt}(T(\cdot)^{-1})$. Performing the derivation of the left hand side of the first equation and multiplying it on the left by $T(\cdot)$ it follows that

$$\begin{aligned} T(\cdot)\overset{\bullet}{\overline{T(\cdot)^{-1}}}z(\cdot) + \frac{d}{dt}z(\cdot) &= T(\cdot)A(\cdot)(T(\cdot)^{-1}z(\cdot)) + T(\cdot)B(\cdot)u(\cdot) \\ y(\cdot) &= C(\cdot)(T(\cdot)^{-1}z(\cdot)) + D(\cdot)u(\cdot). \end{aligned}$$

Then, rearranging terms, the above equation can be rewritten as

$$\begin{aligned} \frac{d}{dt}z(\cdot) &= (T(\cdot)A(\cdot)T(\cdot)^{-1} - T(\cdot)\overset{\bullet}{\overline{T(\cdot)^{-1}}})z(\cdot) + (T(\cdot)B(\cdot))u(\cdot) \\ y(\cdot) &= (C(\cdot)T(\cdot)^{-1})z(\cdot) + D(\cdot)u(\cdot). \end{aligned}$$

It follows from such equality that the new matrices satisfying an i-s-o representation of Σ with the new state variable $z(\cdot)$ are

$$\begin{aligned} A_z(\cdot) &:= T(\cdot)A(\cdot)T(\cdot)^{-1} - T(\cdot)\overset{\bullet}{\overline{T(\cdot)^{-1}}} \\ B_z(\cdot) &:= T(\cdot)B(\cdot) \\ C_z(\cdot) &:= C(\cdot)T(\cdot)^{-1} \\ D_z(\cdot) &:= D(\cdot). \end{aligned}$$

It follows from similar computations for the adjoint system Σ' with the new dual state variable $z'(\cdot) := T'(\cdot)x'(\cdot)$ that

$$\begin{aligned}\frac{d}{dt}z'(\cdot) &= (-T'(\cdot)A(\cdot)^\top T'(\cdot)^{-1} - T'(\cdot)\overset{\bullet}{\overline{T'(\cdot)^{-1}}})z'(\cdot) + (T'(\cdot)C(\cdot)^\top)u'(\cdot)\\ y'(\cdot) &= (B(\cdot)^\top T'(\cdot)^{-1})z'(\cdot) - D(\cdot)^\top u'(\cdot).\end{aligned}$$

Then, the matrices satisfying an i-s-o representation of the adjoint Σ' with the new dual state variable $z'(\cdot)$ are

$$\begin{aligned}A'_{z'}(\cdot) &:= -T'(\cdot)A(\cdot)^\top T'(\cdot)^{-1} - T'(\cdot)\overset{\bullet}{\overline{T'(\cdot)^{-1}}}\\ B'_{z'}(\cdot) &:= T'(\cdot)C(\cdot)^\top\\ C'_{z'}(\cdot) &:= B(\cdot)^\top T'(\cdot)^{-1}\\ D'_{z'}(\cdot) &:= -D(\cdot)^\top.\end{aligned}$$

To prove the second part of the claim let $T'(\cdot) = T(\cdot)^{-\top}$, then the system matrices of the adjoint are given by

$$\begin{aligned}A'_{z'}(\cdot) &:= -(T(\cdot)A(\cdot)T(\cdot)^{-1} + \overset{\bullet}{T}(\cdot)T(\cdot)^{-1})^\top\\ B'_{z'}(\cdot) &:= (C(\cdot)T(\cdot)^{-1})^\top\\ C'_{z'}(\cdot) &:= (T(\cdot)B(\cdot))^\top\\ D'_{z'}(\cdot) &:= -D(\cdot)^\top.\end{aligned}$$

Since $T(\cdot)T(\cdot)^{-1} = I_n$ then

$$\overset{\bullet}{\overline{(T(\cdot)T(\cdot)^{-1})}} = \overset{\bullet}{T}(\cdot)T(\cdot)^{-1} + T(\cdot)\overset{\bullet}{\overline{T(\cdot)^{-1}}} = 0\ ,$$

and consequently

$$\overset{\bullet}{T}(\cdot)T(\cdot)^{-1} = -T(\cdot)\overset{\bullet}{\overline{T(\cdot)^{-1}}}.$$

It follows from Cor. 4.5 that the i-s-o representations of the dual and its adjoint with the new state variables are matched. This concludes the proof. □

Now based on the above definition of the adjoint we introduce the definition of a *self-adjoint* system in the following section.

4.2 Self-adjointness

We define a self-adjoint system as follows.

Definition 4.8. (Crouch et al. (1995); Van der Schaft (1991)) Assume $m = p$; an LTV i-s-o system is *self-adjoint* if there exists $Q(\cdot) = Q(\cdot)^\top$ nonsingular everywhere on $\mathbb{R}$, such that for all pairs (x_k, u_k, y_k), $k = 1, 2$, of i-s-o trajectories it holds that

$$y_1^\top u_2 + u_1^\top y_2 = \frac{d}{dt}\left(x_1^\top Q x_2\right). \tag{4.8}$$

The above definition gives rise to self-adjoint systems which are systems whose input-output trajectories of its primal coincide with that of the dual system. It follows from Cor. 4.5 that the matrices of a matched i-s-o representation of a self-adjoint system satisfy the equations

$$\begin{aligned} A(\cdot) + A(\cdot)^\top &= 0_{n\times n} \\ B(\cdot) - C(\cdot)^\top &= 0_{n\times m} \\ D(\cdot) + D(\cdot)^\top &= 0_{m\times m}\,. \end{aligned} \tag{4.9}$$

The following result gives sufficient conditions for the existence of a matched i-s-o representation for a self-adjoint system.

Proposition 4.9. *Let Σ be a trim self-adjoint system with i-s-o representation induced by $(A(\cdot), B(\cdot), C(\cdot), D(\cdot))$. Assume that there exists a matrix $F(\cdot) \in \mathcal{A}^{n\times n}$ nonsingular everywhere on $\mathbb{R}$ such that for all $t \in \mathbb{R}$ $F(\cdot)^\top F(\cdot) = Q(\cdot)$. Then the i-s-o representation induced by*

$$\begin{aligned} A'(\cdot) &:= F(\cdot)A(\cdot)F(\cdot)^{-1} + \dot{F}(\cdot)F(\cdot)^{-1} \\ B'(\cdot) &:= F(\cdot)B(\cdot) \\ C'(\cdot) &:= C(\cdot)F(\cdot)^{-1} \\ D'(\cdot) &:= D(\cdot)\,, \end{aligned} \tag{4.10}$$

is matched.

Proof. Rewrite (4.8) using the i-s-o equations of Σ

$$\begin{aligned} & (C(\cdot)x_1(\cdot) + D(\cdot)u_1(\cdot))^\top u_2(\cdot) + u_1^\top (C(\cdot)x_2(\cdot) + D(\cdot)u_2(\cdot)) \\ = \;& (A(\cdot)x_1(\cdot) + B(\cdot)u_1(\cdot))^\top Q(\cdot)x_2(\cdot) + x_1(\cdot)^\top \dot{Q}(\cdot)x_2(\cdot) \\ & + x_1(\cdot)^\top Q(\cdot)(A(\cdot)x_2(\cdot) + B(\cdot)u_2(\cdot)). \end{aligned}$$

It is matter of straightforward verification that such equality can also be written as

$$\begin{bmatrix} x_1(\cdot) \\ u_1(\cdot) \end{bmatrix}^\top \begin{bmatrix} A(\cdot)^\top Q(\cdot) + Q(\cdot)A(\cdot) + \dot{Q}(\cdot) & Q(\cdot)B(\cdot) - C(\cdot)^\top \\ B(\cdot)^\top Q(\cdot) - C(\cdot) & -D(\cdot) - D(\cdot)^\top \end{bmatrix} \begin{bmatrix} x_2(\cdot) \\ u_2(\cdot) \end{bmatrix} = 0\,. \tag{4.11}$$

It follows that for every pair of vectors $v_1, v_2 \in \mathbb{R}^{n+m}$ there exist $t \in \mathbb{R}$ and trajectories $\mathrm{col}(x_1, u_1)$ and $\mathrm{col}(x_2, u_2)$ whose value at t is v_1, respectively v_2. Then it follows that the central matrix of (4.11) is equal to zero.

Now since $Q(\cdot) = F(\cdot)^\top F(\cdot)$ it follows that the $(1,1)$-block of the central matrix is given by

$$A(\cdot)^\top F(\cdot)^\top F(\cdot) + F(\cdot)^\top F(\cdot) A(\cdot) + \overset{\bullet}{\overline{(F(\cdot)^\top F(\cdot))}} = 0_{n\times n}.$$

Since $\overset{\bullet}{\overline{(F(\cdot)^\top F(\cdot))}} = F(\cdot)^\top \overset{\bullet}{F}(\cdot) + \overset{\bullet}{F}(\cdot)^\top F(\cdot)$ and multiplying on the left by $F(\cdot)^{-\top}$ and on the right by $F(\cdot)^{-1}$ in the above equation it follows that

$$F(\cdot)A(\cdot)F(\cdot)^{-1} + \overset{\bullet}{F}(\cdot)F(\cdot)^{-1} + F(\cdot)^{-\top}A(\cdot)^\top F(\cdot)^\top + F(\cdot)^{-\top}\overset{\bullet}{F}(\cdot)^\top = \quad 0_{n\times n}. \tag{4.12}$$

Now define

$$A'(\cdot) := F(\cdot)A(\cdot)F(\cdot)^{-1} + \overset{\bullet}{F}(\cdot)F(\cdot)^{-1},$$

then eq. (4.12) can be written as

$$A'(\cdot) + A'(\cdot)^\top = 0_{n\times n}.$$

Following similar computation for the $(1,2)$-block of the central matrix in (4.11) it follows that

$$F(\cdot)^\top F(\cdot)B(\cdot) - C(\cdot)^\top \quad = \quad 0_{n\times m},$$

and multiplying on the left by $F(\cdot)^{-\top}$ we obtain

$$F(\cdot)B(\cdot) - F(\cdot)^{-\top}C(\cdot)^\top = 0_{n\times m}.$$

From the above equation, define

$$\begin{aligned} B'(\cdot) &:= F(\cdot)B(\cdot) \\ C'(\cdot) &:= C(\cdot)F(\cdot)^{-1}, \end{aligned}$$

then it follows from such equations that

$$B'(\cdot) - C'(\cdot)^\top = 0_{n\times m}.$$

Note also that from the $(2,2)$-block of (4.11) we can define

$$D'(\cdot) := D(\cdot).$$

We conclude that the conditions for self-adjointness are:

$$\begin{aligned} F(\cdot)A(\cdot)F(\cdot)^{-1} + \dot{F}(\cdot)F(\cdot)^{-1} + \left(F(\cdot)A(\cdot)F(\cdot)^{-1} + \dot{F}(\cdot)F(\cdot)^{-1}\right)^{\top} &= 0_{n\times n} \\ F(\cdot)B(\cdot) - \left(C(\cdot)F(\cdot)^{-1}\right)^{\top} &= 0_{n\times m} \\ D(\cdot) + D(\cdot)^{\top} &= 0_{m\times m}, \end{aligned}$$

or using $A'(\cdot)$, $B'(\cdot)$, $C'(\cdot)$ and $D'(\cdot)$ as defined previously in this proof:

$$\begin{aligned} A'(\cdot) + A'(\cdot)^{\top} &= 0_{n\times n} \\ B'(\cdot) - C'(\cdot)^{\top} &= 0_{n\times m} \\ D'(\cdot) + D'(\cdot)^{\top} &= 0_{m\times m}. \end{aligned} \tag{4.13}$$

It follows from (4.13) and Cor. 4.5 that the i-s-o representation of a self-adjoint system induced by $(A'(\cdot), B'(\cdot), C'(\cdot), D'(\cdot))$ is matched. This concludes the proof. □

4.3 Summary

In this chapter we introduced a trajectory-based definition of duality. We defined how external properties at the level of inputs and outputs are reflected in the internal ones at the level of the state. We also showed how the internal structure is affected when the basis of the state space changes. Self-adjoint property was introduced.

Chapter 5

Duality approach for system identification

In this chapter, we introduce the duality-based approach for the identification of LTV systems. We use the fact that the internal information at the level of the state is related to the external structures at the level of inputs and output. In this framework, we present two algorithms. The first one assumes that the dual trajectories are available for experiments. In the second one, we consider the case when the system is self-adjoint, and consequently the trajectories of the primal are also trajectories of the dual.

5.1 Duality as a system identification approach

In this section we propose a procedure to identify an LTV system using duality. We state the problem of this section as follows. Consider the following definition of an *unfalsified model*.

Definition 5.1. Let Σ_{ss} be a LTV system. Assume that from N number of input-output trajectories of Σ_{ss} (i.e. $(u_i(\cdot), y_i(\cdot))_{i=1,..,N}$), we compute its corresponding state trajectories $x_i(\cdot)$. Then the matrices $A(\cdot), B(\cdot), C(\cdot), D(\cdot)$ correspond to an unfalsified model for the trajectories $(u_i(\cdot), y_i(\cdot))$ if they satisfy

$$\begin{bmatrix} \frac{d}{dt}x_1(\cdot) & \frac{d}{dt}x_2(\cdot) & \cdots & \frac{d}{dt}x_N(\cdot) \\ y_1(\cdot) & y_2(\cdot) & \cdots & y_N(\cdot) \end{bmatrix} = \begin{bmatrix} A(\cdot) & B(\cdot) \\ C(\cdot) & D(\cdot) \end{bmatrix} \begin{bmatrix} x_1(\cdot) & x_2(\cdot) & \cdots & x_N(\cdot) \\ u_1(\cdot) & u_2(\cdot) & \cdots & u_N(\cdot) \end{bmatrix}.$$

Identification problem.

Given

- a set of N input-output trajectories of the primal Σ_s

$$\{(u_i(\cdot), y_i(\cdot)) \mid u_i(t) \in \mathbb{R}^m,\ y_i(t) \in \mathbb{R}^p,\ x_i(0) = 0_{n\times 1}\}_{i=1,..,N},$$

- and its dual Σ_a

$$\{(u_i'(\cdot), y_i'(\cdot)) \mid u_i'(t) \in \mathbb{R}^p,\ y_i'(t) \in \mathbb{R}^m,\ x_i'(0) = 0_{n\times 1}\}_{i=1,..,N}$$

generated by a LTV system.

Compute

- matrices $A(\cdot), B(\cdot), C(\cdot), D(\cdot)$ corresponding to an unfalsified model for $(u_i(\cdot), y_i(\cdot))_{i=1,..,N}$

In order to solve this problem we use duality to exploit the fact that external properties, i.e. inputs and outputs, are reflected in the internal ones, i.e. states. In the first stage of the identification procedure we develop a procedure to obtain state trajectories. Then system matrices of an unfalsified model for the data are computed. Theses stages will be explained in more detail in the following sections.

5.2 State construction

In the following we develop a procedure to construct the state trajectories of a state space representation associated with the data.

Let eq. (4.1), then integrating both sides of (4.1) in $[0, t]$ we obtain

$$\int_0^t u(\tau)^\top y'(\tau) + y(\tau)^\top u'(\tau)\ d\tau = x(t)^\top Q(t)x'(t) - x(0)^\top Q(0)x'(0).$$

Using the assumption that $x_k(0) = x_k'(0) = 0$ for $k = 1, \ldots, N$, it follows that

$$\int_0^t u(\tau)^\top y'(\tau) + y(\tau)^\top u'(\tau)\ d\tau = x(t)^\top Q(t)x'(t). \tag{5.1}$$

Now consider N input-output trajectories of the primal, $(u_k, y_k)_{k=1,..,N}$, and its dual, $(u'_k, y'_k)_{k=1,..,N}$. Define a matrix $E(t) \in \mathbb{R}^{N\times N}$ whose $(k,l) - th$ entries are given by

$$e_{k,\ell}(t) := \int_0^t u_k(\tau)^\top y'_\ell(\tau) + y_k(\tau)^\top u'_\ell(\tau)\ d\tau \quad k,\ell = 1,..,N. \tag{5.2}$$

Note that (5.2) defines the value of $E(\cdot)$ at t, and $E(\cdot)$ is the corresponding function. Then it follows from such definition and (5.1) that

$$E(\cdot) = \begin{bmatrix} x_1(\cdot)^\top \\ \vdots \\ x_N(\cdot)^\top \end{bmatrix} Q(\cdot) \begin{bmatrix} x'_1(\cdot) & \dots & x'_N(\cdot) \end{bmatrix}. \tag{5.3}$$

The following definition is crucial for the main result of this section.

Definition 5.2. (Heij (1989)) Let $L \subseteq \mathbb{R}^d$ be a linear space. Then a set $S \subset L$ is generic in L if there exists a *proper algebraic variety*[1] $\mathcal{V}$ in L such that $S \supset (L \backslash \mathcal{V})$.

An example of *genericity* is illustrated in the following. Let A be a real square matrix of dimension $n \times n$ whose (i,j)-th element is denoted by a_{ij}. Then a proper algebraic variety in $\mathbb{R}^{n^2}$ is defined by

$$\mathcal{V} := \{\text{col}(a_{11},..,a_{nn}) \in \mathbb{R}^{n^2} \mid \det(A) = 0\}. \tag{5.4}$$

Then the generic set $S \supset (\mathbb{R}^{n^2} \backslash \mathcal{V})$ is defined by all possible elements on $\mathbb{R}^{n^2}$ such that A is nonsingular, i.e.

$$S := \{\text{col}(a_{11},..,a_{nn}) \in \mathbb{R}^{n^2} \mid \det(A) \neq 0\}. \tag{5.5}$$

It follows from the above example and the definition of genericity that a matrix $A \in \mathbb{R}^{n\times n}$, whose entries belong to the generic set, is generically nonsingular.

Equation (5.3) and Def. 5.2 leads us to the first result of this section.

Proposition 5.3. *Assume $(A(\cdot), B(\cdot))$ are controllable in the sense of Theorem 3.3. Then, generically*

$$rank \begin{bmatrix} x_1(\cdot) & \cdots & x_N(\cdot) \\ u_1(\cdot) & \cdots & u_N(\cdot) \end{bmatrix} = n + m \tag{5.6}$$

Proof. Consider the matrix of functions formed by the state trajectories $\{x_k(\cdot)\}_{k=1,..,N}$ and input trajectories $\{u_k(\cdot)\}_{k=1,..,N}$:

$$F(\cdot) := \begin{bmatrix} x_1(\cdot) & \cdots & x_N(\cdot) \\ u_1(\cdot) & \cdots & u_N(\cdot) \end{bmatrix} \tag{5.7}$$

[1] See Definition A.4 in Appendix A.

First, we use a variable transformation for the state space representation of Σ_s to simplify the state equations, see Def. 3.1. Every property defined in the transformed system is valid for the original one.

Consider the matrices $A(\cdot), B(\cdot), C(\cdot)$ and $D(\cdot)$ satisfying Σ_s, and let a nonsingular matrix $T(\cdot) \in \mathcal{A}^{n\times n}$ be a matrix such that for all $t \in \mathbb{R}$

$$\dot{T}(t) = A(t)T(t)$$

is satisfied. From Silverman and Meadows (1967), a transition matrix for the system Σ_s is defined as

$$\phi(t,\tau) = T(t)T(\tau)^{-1}.$$

Now let $\phi(\tau,t) := T(\tau)T(t)^{-1}$ be the inverse of the transition matrix, then the following state variable transformation is given

$$z(t) = \phi(\tau,t)x(t). \tag{5.8}$$

Now we apply the variable transformation to the system Σ_s using (5.8). From (5.8) we have that $x(t) = \phi(t,\tau)z(t)$. Substituting $x(t)$ with $\phi(t,\tau)z(t)$ in Σ_s for a fixed but otherwise arbitrary τ we conclude

$$\begin{aligned}\frac{d}{dt}(\phi(\cdot,\tau)z(\cdot)) &= A(\cdot)(\phi(\cdot,\tau)z(\cdot)) + B(\cdot)u(\cdot)\\ y(\cdot) &= C(\cdot)(\phi(\cdot,\tau)z(\cdot)) + D(\cdot)u(\cdot).\end{aligned}$$

Differentiating the left hand side of the above equation we obtain

$$\begin{aligned}\left(\frac{d}{dt}\phi(\cdot,\tau)\right)z(\cdot) + \phi(\cdot,\tau)\left(\frac{d}{dt}z(\cdot)\right) &= A(\cdot)\phi(\cdot,\tau)z(\cdot) + B(\cdot)u(\cdot)\\ y(\cdot) &= C(\cdot)\phi(\cdot,\tau)z(\cdot) + D(\cdot)u(\cdot),\end{aligned}$$

where $\frac{d}{dt}\phi(t,\tau)$ is given by

$$\begin{aligned}\frac{d}{dt}\phi(t,\tau) &= \frac{d}{dt}(T(t)T(\tau)^{-1})\\ &= \dot{T}(t)T(\tau)^{-1}\\ &= A(t)T(t)T(\tau)^{-1}\\ &= A(t)\phi(t,\tau).\end{aligned}$$

Then, it follows that the transformed system with the new state variable $z(\cdot)$ is represented by

$$\begin{aligned}\frac{d}{dt}z(\cdot) &= \phi(\tau,\cdot)B(\cdot)u(\cdot)\\ y(\cdot) &= C(\cdot)\phi(\cdot,\tau)z(\cdot) + D(\cdot)u(\cdot).\end{aligned} \tag{5.9}$$

Now consider the Wronskian[2] of the matrix (5.7) with the new state variable $z(\cdot)$. Let us denote $z(\cdot)^{(n)}$ as the n-th derivative of $z(\cdot)$, then the Wronskian of (5.7) with the new state variable $z(\cdot)$ is defined by

$$W_i(\cdot) := \begin{bmatrix} z_i(\cdot) & \cdots & z_i(\cdot)^{(n+m-1)} \\ u_i(\cdot) & \cdots & u_i(\cdot)^{(n+m-1)} \end{bmatrix}_{i=1,..,N}. \tag{5.10}$$

Since

$$\begin{aligned} z(\cdot)^{(1)} &= \phi(\tau,\cdot)B(\cdot)u(\cdot) \\ z(\cdot)^{(2)} &= \phi(\tau,\cdot)\begin{bmatrix} B(\cdot) & -A(\cdot)B(\cdot) + \dot{B}(\cdot) \end{bmatrix}\begin{bmatrix} u(\cdot)^{(1)} \\ u(\cdot) \end{bmatrix} \\ &\vdots \\ z(\cdot)^{(\ell)} &= \phi(\tau,\cdot)\begin{bmatrix} P_1(\cdot) & P_2(\cdot) & \cdots & P_\ell(\cdot) \end{bmatrix}\begin{bmatrix} u(\cdot)^{(\ell-1)} \\ u(\cdot)^{(\ell-2)} \\ \vdots \\ u(\cdot) \end{bmatrix} \end{aligned}$$

where

$$P_0(\cdot) := B(\cdot) \quad P_{s+1} := -A(\cdot)P_s(\cdot) + \dot{P}_s(\cdot) \quad s = 0,..,\ell-2,$$

we can express $W_i(\cdot)$ as follows. Let

$$\mathcal{C}_i(\cdot) := \begin{bmatrix} z_i(\cdot) & \phi(\tau,\cdot)\begin{bmatrix} P_1(\cdot) & P_2(\cdot) & \cdots & P_{n+m-1}(\cdot) \end{bmatrix} \end{bmatrix}_{i=1,...,N}$$

and

$$\mathcal{U}_{1,i}(\cdot) := \begin{bmatrix} I & 0 & 0 & \cdots & \cdots & 0 \\ 0 & u_i(\cdot) & u_i(\cdot)^{(1)} & \cdots & \cdots & u_i(\cdot)^{(n+m-2)} \\ 0 & 0 & u_i(\cdot) & \ddots & \cdots & u_i(\cdot)^{(n+m-3)} \\ 0 & 0 & 0 & \ddots & \ddots & \vdots \\ \vdots & \vdots & \vdots & \ddots & \ddots & u_i(\cdot)^{(1)} \\ 0 & 0 & 0 & \cdots & 0 & u_i(\cdot) \end{bmatrix}_{i=1,...,N}$$

$$\mathcal{U}_{2,i}(\cdot) := \begin{bmatrix} u_i(\cdot) & \cdots & u_i(\cdot)^{(n+m-1)} \end{bmatrix}_{i=1,...,N}.$$

Then the Wronskian of $F(\cdot)$ can be written as

$$W_i(\cdot) = \begin{bmatrix} \mathcal{C}_i(\cdot) & 0 \\ 0 & I \end{bmatrix}\begin{bmatrix} \mathcal{U}_{1,i}(\cdot) \\ \mathcal{U}_{2,i}(\cdot) \end{bmatrix}.$$

[2] See Definition A.5 in Appendix A.

According to lemma 1 in p.67 of Silverman and Meadows (1967), if $\text{rank}(W_i(\cdot)) = n+m$ on $[t_0, t_1]$, then the rows of $F(\cdot)$ are *linearly independent*[3] as a set of functions on the interval $[t_0, t_1]$.

By contradiction, assume that the rank of $W_i(\cdot)$ is not $(n+m)$, then every $(n+m)$-dimensional submatrix of $W_i(\cdot)$ is singular, i.e. every $(n+m)$-dimensional submatrix has $(n+m)$-column/row function vectors that are linearly dependent. Consequently, there exists a non-all zero vector $[\alpha_1(\cdot)^\top \ \alpha_2(\cdot)^\top]$ such that

$$\begin{bmatrix}\alpha_1(\cdot)^\top & \alpha_2(\cdot)^\top\end{bmatrix}\begin{bmatrix}\mathcal{C}_i(\cdot) & 0\\ 0 & I\end{bmatrix}\begin{bmatrix}\mathcal{U}_{1,i}(\cdot)\\ \mathcal{U}_{2,i}(\cdot)\end{bmatrix} = 0,$$

then

$$\alpha_1^\top \mathcal{C}_i(\cdot)\mathcal{U}_{1,i}(\cdot) + \alpha_2^\top \mathcal{U}_{2,i}(\cdot) = 0.$$

It follows from genericity that $\alpha_2^\top \mathcal{U}_{2,i}(\cdot) = 0$ iff $\alpha_2 \top = 0$. Then

$$0 = \alpha_1^\top \mathcal{C}_i(\cdot)\mathcal{U}_{1,i}(\cdot) \tag{5.11}$$

with a non-all zero vector $\alpha_1^\top$.

Note that the Wronskian of $\phi(\tau, \cdot)B(\cdot)$ can be defined as

$$W_z(\cdot) := \phi(\tau, \cdot)C_n,$$

see Section 3.1.2.1. Then $\mathcal{C}_i(\cdot)$ in (5.11) can be written as

$$\mathcal{C}_i(\cdot) = \begin{bmatrix} z_i(\cdot) & W_z(\cdot) & \phi(\tau, \cdot)\begin{bmatrix}P_n(\cdot) & P_2(\cdot) & \cdots & P_{n+m-1}(\cdot)\end{bmatrix}\end{bmatrix}.$$

It follows from Theorem 3.3 and eq. (5.9) that if the system Σ_s is controllable, then the rows of $W_z(\cdot)$ are linearly independent. This implies that the rank of $\mathcal{C}_i(\cdot)$ is at least n, and consequently full row rank.

Since the main diagonal of $\mathcal{U}_{1,i}(\cdot)$ is clearly nonzero, it follows that $\mathcal{U}_{1,i}$ has full rank. Consequently $\text{rank}(\mathcal{C}_i(\cdot)\mathcal{U}_{1,i}(\cdot))$ =$\text{rank}(\mathcal{C}_i(\cdot)) = n$ which contradicts (5.11).

It follows that $\text{rank}(W_i(\cdot)) = m + n$ and consequently the rows of $F(\cdot)$ are linearly independent. Then

$$\text{rank}\begin{bmatrix}x_1(\cdot) & \cdots & x_N(\cdot)\\ u_1(\cdot) & \cdots & u_N(\cdot)\end{bmatrix} = m + n$$

is generically true. This concludes the proof. □

[3] See Definition A.1 in Appendix A,

From Prop. 5.3, we can conclude that generically the rank of the matrix $E(\cdot)$ whose entries are defined by (5.3) equals the dimension of the state. Now let n be the state dimension, and consider the partition

$$E(\cdot) = \begin{bmatrix} E_{1,1}(\cdot) & E_{1,2}(\cdot) \\ E_{2,1}(\cdot) & E_{2,2}(\cdot) \end{bmatrix} \tag{5.12}$$

where $E_{1,1}(\cdot) \in \mathcal{A}^{n\times n}$, $E_{1,2}(\cdot) \in \mathcal{A}^{n\times(N-n)}$, $E_{2,1}(\cdot) \in \mathcal{A}^{(N-n)\times n}$ and $E_{2,2}(\cdot) \in \mathcal{A}^{(N-n)\times(N-n)}$. Then the following proposition holds.

Proposition 5.4. *Let $E(\cdot) \in \mathcal{A}^{N\times N}$ be a matrix of functions whose $(k,l)-th$ entries at $t \in \mathbb{R}$ are defined by (5.3). Then the $n\times n$ $(1,1)-$block of $E(\cdot)$ is generically non-singular.*

Proof. From Prop. 5.3, generically $\text{rank}(E(\cdot)) = n$. It follows that the number of linearly independent rows and columns of $E(\cdot)$ is n.

Now let the partition in (5.12) be given. It follows from Def. 5.2 that a proper algebraic variety $\mathcal{V}$ in $\mathcal{A}^{n^2}$ is defined by the $(k,l)-th$ entries of $E_{1,1}(\cdot)$ such that $\det(E_{1,1}(\cdot)) = 0$. Then generically the elements of $E_{1,1}(\cdot)$ belongs to the generic set $S \supset (\mathcal{A}^{n^2}\backslash\mathcal{V})$. This concludes the proof. □

The following result produces a factorization of the state trajectories.

Proposition 5.5. *Let $E(\cdot) \in \mathcal{A}^{N\times N}$ and the conditions of Prop. 5.4 be given. Then generically*

$$E(\cdot) := \begin{bmatrix} I_n \\ E_{2,1}(\cdot)E_{1,1}(\cdot)^{-1} \end{bmatrix} E_{1,1}(\cdot) \begin{bmatrix} I_n & E_{1,1}(\cdot)^{-1}E_{1,2}(\cdot) \end{bmatrix} \tag{5.13}$$

holds.

Proof. From Prop. 5.4 $E_{1,1}(\cdot)$ is generically nonsingular. It follows that (5.13) can be decomposed as

$$E(\cdot) = \begin{bmatrix} I_n & 0_{n\times(N-n)} \\ E_{2,1}(\cdot)E_{1,1}(\cdot)^{-1} & I_{(N-n)} \end{bmatrix} \begin{bmatrix} E_{1,1}(\cdot) & 0_{n\times(N-n)} \\ 0_{(N-n)\times n} & \triangle(\cdot) \end{bmatrix} \begin{bmatrix} I_n & E_{1,1}(\cdot)^{-1}E_{1,2}(\cdot) \\ 0_{(N-n)\times n} & I_{(N-n)} \end{bmatrix}$$

where $\triangle(\cdot) := E_{2,2}(\cdot) - E_{2,1}(\cdot)E_{1,1}(\cdot)^{-1}E_{1,2}(\cdot)$. It follows from Prop. 5.3 that $\text{rank}(E(\cdot)) = n$, then $\text{rank}(E(\cdot)) = \text{rank}(E_{1,1}(\cdot))$ implies $\triangle(\cdot) = 0$. This concludes the proof. □

We show in Prop. 5.5 a procedure to factorize the matrix $E(\cdot)$ generated in (5.3). In the following we show that this factorization corresponds to the factorization of state trajectories of the to be identified LTV system.

Proposition 5.6. *Let $E(\cdot) \in \mathcal{A}^{N\times N}$ be defined by (5.2), and let the decomposition of $E(\cdot)$ be given as in Prop. 5.5. Define*

$$X(\cdot) := \begin{bmatrix} I_n & E_{1,1}(\cdot)^{-\top} E_{2,1}(\cdot)^{\top} \end{bmatrix} \in \mathcal{A}^{n\times N}, \tag{5.14}$$

then the matrices $(A(\cdot), B(\cdot), C(\cdot), D(\cdot))$ define an unfalsified model if they satisfy

$$\begin{bmatrix} \dot{X}(\cdot) \\ Y(\cdot) \end{bmatrix} = \begin{bmatrix} A(\cdot) & B(\cdot) \\ C(\cdot) & D(\cdot) \end{bmatrix} \begin{bmatrix} X(\cdot) \\ U(\cdot) \end{bmatrix}. \tag{5.15}$$

Proof. From Prop. 5.3, the matrix

$$\begin{bmatrix} X(\cdot) \\ U(\cdot) \end{bmatrix}$$

has full row rank. Then it admits a *right inverse*[4] $R(\cdot)$ such that

$$\begin{bmatrix} X(\cdot) \\ U(\cdot) \end{bmatrix} R(\cdot) = I_{n+m}.$$

It follows that the matrices $(A(\cdot), B(\cdot), C(\cdot), D(\cdot))$ can be computed as

$$\begin{bmatrix} \dot{X}(\cdot) \\ Y(\cdot) \end{bmatrix} R(\cdot) = \begin{bmatrix} A(\cdot) & B(\cdot) \\ C(\cdot) & D(\cdot) \end{bmatrix}. \tag{5.16}$$

Since the sums, products, and compositions of analytic functions are analytic, then the entries of the resulting system matrices in (5.16) are also analytic. This yields the claim. □

It follows that the state trajectories $X(\cdot)$ defined in Prop. 5.6 satisfy an i-s-o representation induced by matrices $(A(\cdot), B(\cdot), C(\cdot), D(\cdot))$ such that the conditions of Prop. 5.6 hold. Note that this factorization is generically valid for all $t \in \mathbb{R}$. In contrast to state factorizations of classical subspace methods, we can choose the basis of the state space by choosing a special factorization.

5.3 Matrix computation

Now that we have a procedure for obtaining the state trajectories from the input-output data, we proceed to the computation of the matrices of the state space representation as follows.

[4] See Definition A.2 in Appendix A.

Computation of system matrices

Given

$$Y(\cdot) := [y_1(\cdot) \ ... \ y_N(\cdot)] \in \mathcal{A}^{m\times N}$$
$$U(\cdot) := [u_1(\cdot) \ ... \ u_N(\cdot)] \in \mathcal{A}^{m\times N}$$
$$X(\cdot) := [x_1(\cdot) \ ... \ x_N(\cdot)] \in \mathcal{A}^{n\times N}.$$

Solve

$$\begin{bmatrix} \dot{X}(\cdot) \\ Y(\cdot) \end{bmatrix} = \begin{bmatrix} A(\cdot) & B(\cdot) \\ C(\cdot) & D(\cdot) \end{bmatrix} \begin{bmatrix} X(\cdot) \\ U(\cdot) \end{bmatrix}$$

for $(A(\cdot), B(\cdot), C(\cdot), D(\cdot))$.

Let the i-s-o representation of the to be identified LTV system be given by

$$\frac{d}{dt}x(\cdot) = A(\cdot)x(\cdot) + B(\cdot)u(\cdot)$$
$$y(\cdot) = C(\cdot)x(\cdot) + D(\cdot)u(\cdot)$$

whose matrix form is

$$\begin{bmatrix} \frac{d}{dt}x(\cdot) \\ y(\cdot) \end{bmatrix} = \begin{bmatrix} A(\cdot) & B(\cdot) \\ C(\cdot) & D(\cdot) \end{bmatrix} \begin{bmatrix} x(\cdot) \\ u(\cdot) \end{bmatrix}.$$

For N number of trajectories, such a system is given by

$$\begin{bmatrix} \frac{d}{dt}x_1(\cdot) & \cdots & \frac{d}{dt}x_N(\cdot) \\ y_1(\cdot) & \cdots & y_N(\cdot) \end{bmatrix} = \begin{bmatrix} A(\cdot) & B(\cdot) \\ C(\cdot) & D(\cdot) \end{bmatrix} \begin{bmatrix} x_1(\cdot) & \cdots & x_N(\cdot) \\ u_1(\cdot) & \cdots & u_N(\cdot) \end{bmatrix}.$$

Then using the input-output trajectories for N number of experiments defined by

$$Y(\cdot) := [y_1(\cdot) \ ... \ y_N(\cdot)]$$
$$U(\cdot) := [u_1(\cdot) \ ... \ u_N(\cdot)],$$

and state trajectories $X(\cdot)$ defined as in Prop. (5.6), matrices $A(\cdot)$,$B(\cdot)$,$C(\cdot)$ and $D(\cdot)$ of a state representation associated with such trajectories can be computed by solving the following set of functional equations

$$\begin{bmatrix} \dot{X}(\cdot) \\ Y(\cdot) \end{bmatrix} = \begin{bmatrix} A(\cdot) & B(\cdot) \\ C(\cdot) & D(\cdot) \end{bmatrix} \begin{bmatrix} X(\cdot) \\ U(\cdot) \end{bmatrix}. \tag{5.17}$$

Let

$$Y(\cdot) := [y_1(\cdot) \ ... \ y_N(\cdot)]$$
$$U(\cdot) := [u_1(\cdot) \ ... \ u_N(\cdot)]$$
$$X(\cdot) := [x_1(\cdot) \ ... \ x_N(\cdot)],$$

and define the derivative of the state trajectories as

$$\overset{\bullet}{X}(\cdot) := \frac{d}{dt}X(\cdot).$$

Then we compute $(A(\cdot), B(\cdot), C(\cdot), D(\cdot))$ as follows. Define

$$M(\cdot) := \begin{bmatrix} X(\cdot) \\ U(\cdot) \end{bmatrix}$$

From prop 5.3, it follows that the matrix $M(\cdot)$ is full row rank, therefore it admits a right inverse $R(\cdot) \in \mathcal{A}^{N\times(n+m)}$ such that

$$M(\cdot)R(\cdot) = I_{n+m}.$$

It follows that Prop. 5.3 gives us sufficient conditions for the existence of a solution of the following linear functional equation,

$$\begin{bmatrix} \overset{\bullet}{X}(\cdot) \\ Y(\cdot) \end{bmatrix} R(\cdot) = \begin{bmatrix} A(\cdot) & B(\cdot) \\ C(\cdot) & D(\cdot) \end{bmatrix}. \tag{5.18}$$

Then, the system matrices $(A(\cdot), B(\cdot), C(\cdot), D(\cdot))$ computed in (5.18) correspond to an unfalsified model for $(u_i(\cdot), y_i(\cdot))_{i=1,..,N}$.

We now introduce the algorithm for LTV system identification in Algorithm 4.

Algorithm 4 Duality-based algorithm

Input: $(u_i(\cdot), y_i(\cdot))$, $(u_i'(\cdot), y_i'(\cdot))$, $i = 1, .., N$
Output: $(A(\cdot), B(\cdot), C(\cdot), D(\cdot))$
Assumptions: Genericity, $n < N$

1. Build $E(\cdot)$ as in (5.2):

$$e_{k,\ell}(t) := \int_0^t u_k(\tau)^\top y_\ell'(\tau) + y_k(\tau)^\top u_\ell'(\tau)\, d\tau \quad k, \ell = 1, .., N,$$
$$E(t) = \begin{bmatrix} e_{1,1}(t) & \cdot & e_{1,N} \\ \vdots & . & \vdots \\ e_{N,1}(t) & \cdot & e_{N,N} \end{bmatrix}.$$

2. Compute the state dimension: $\operatorname{rank}(E(\cdot)) = n$.

3. Partition $E(\cdot)$ as
$$E(\cdot) := \begin{bmatrix} E_{1,1}(\cdot) & E_{1,2}(\cdot) \\ E_{2,1}(\cdot) & E_{2,2}(\cdot) \end{bmatrix}$$
such that $E_{1,1}(\cdot) \in \mathcal{A}^{n \times n}$ is generically nonsingular.

4. Define $X(\cdot) := [I_n \;\; E_{1,1}(\cdot)^{-\top} E_{2,1}(\cdot)^\top]$, $Y(\cdot) := [y_1(\cdot) \;\; \cdots \;\; y_N(\cdot)]$ and $U(\cdot) := [u_1(\cdot) \;\; \cdots \;\; u_N(\cdot)]$.

5. Compute $\overset{\bullet}{X}$.

6. Compute $R(\cdot) \in \mathcal{A}^{N \times (n+m)}$ such that
$$\begin{bmatrix} X(\cdot) \\ U(\cdot) \end{bmatrix} R(\cdot) = I_{n+m}.$$

7. Define $(A(\cdot), B(\cdot), C(\cdot), D(\cdot))$ from (5.18).

5.4 The self-adjoint case

In the previous sections we discussed the case when one has the primal and the dual system available for experiments. If the LTV system is self-adjoint, the input-output map of the primal and dual coincide. Then every input-output trajectory of the primal is also a trajectory of the dual. In this context, the identification problem for self-adjoint systems is stated as follows.

Identification problem. The self-adjoint case.

Given

- a set of N input-output trajectories

$$\{(u_i(\cdot), y_i(\cdot)) \mid u_i(t) \in \mathbb{R}^m,\ y_i(t) \in \mathbb{R}^m,\ x_i(0) = 0_{n\times 1}\}_{i=1,..,N},$$

generated by a LTV self-adjoint system.

Compute

- matrices $A(\cdot), B(\cdot), C(\cdot), D(\cdot)$ corresponding to an unfalsified model for $(u_i(\cdot), y_i(\cdot))_{i=1,..,N}$.

Since the LTV system is self-adjoint we use (4.8) for our procedure.

Using the assumption that $x_k(0) = 0$ for $k = 1, \ldots, N$, we have that integrating both sides of (4.8) in $[0, t]$ we can obtain

$$\int_0^t u(\tau)^\top y(\tau) + y(\tau)^\top u_(\tau)\ d\tau = x(t)^\top Q(t) x(t). \tag{5.19}$$

We now define a matrix $E(t) \in \mathbb{R}^{N\times N}$ whose $(k, l) - th$ entries are given by

$$e_{k,\ell}(t) := \int_0^t u_k(\tau)^\top y_\ell(\tau) + y_k(\tau)^\top u_\ell(\tau)\ d\tau \quad k, \ell = 1, .., N. \tag{5.20}$$

It follows from (5.19) that

$$E(\cdot) = \begin{bmatrix} x_1(\cdot)^\top \\ \vdots \\ x_N(\cdot)^\top \end{bmatrix} Q(\cdot) \begin{bmatrix} x_1(\cdot) & \ldots & x_N(\cdot) \end{bmatrix}. \tag{5.21}$$

Note that $E(\cdot) \in \mathbb{R}^{N\times N}$ defined in (5.21) is now a symmetric matrix. This is from the fact that we can choose the input-output trajectories of dual and the primal to be the same for every $i - th$ experiment due to the self-adjointness property.

From Prop. 5.3, we have that generically the rank of the matrix $E(\cdot)$ whose entries are defined by (5.20) equals the dimension of the state, i.e. $\text{rank}(E(\cdot)) = n$. Now consider the partition

$$E(\cdot) = \begin{bmatrix} E_{1,1}(\cdot) & E_{1,2}(\cdot) \\ E_{1,2}(\cdot)^\top & E_{2,2}(\cdot) \end{bmatrix} \tag{5.22}$$

where $E_{1,1}(\cdot) \in \mathcal{A}^{n\times n}$, $E_{1,2}(\cdot) \in \mathcal{A}^{n\times(N-n)}$, and $E_{2,2}(\cdot) \in \mathcal{A}^{(N-n)\times(N-n)}$. It follows from 5.4 that $E_{1,1}(\cdot)$ in (6.23) is nonsingular. Then the following proposition holds.

Proposition 5.7. *Let $E(\cdot) \in \mathcal{A}^{N\times N}$ be defined as in (5.20). Then*

$$E(\cdot) := \begin{bmatrix} I_n \\ E_{1,2}(\cdot)^\top E_{1,1}(\cdot)^{-\top} \end{bmatrix} E_{1,1}(\cdot) \begin{bmatrix} I_n & E_{1,1}(\cdot)^{-1}E_{1,2}(\cdot) \end{bmatrix} \tag{5.23}$$

holds.

Proof. Follows from the proof of Prop 5.5. □

Now we prove that this factorization corresponds to the factorization of state trajectories of the to be identified LTV self-adjoint system as follows.

Proposition 5.8. *Let $E(\cdot) \in \mathcal{A}^{N\times N}$ be defined by (5.20), and let the decomposition of $E(\cdot)$ be given as in Prop. 5.7. Define*

$$X(\cdot) := \begin{bmatrix} I_n & E_{1,1}(\cdot)^{-1}E_{1,2}(\cdot) \end{bmatrix} \in \mathcal{A}^{n\times N}. \tag{5.24}$$

Then the matrices $(A(\cdot), B(\cdot), C(\cdot), D(\cdot))$ defines an unfalsified model if they satisfy

$$\begin{bmatrix} \dot{X}(\cdot) \\ Y(\cdot) \end{bmatrix} = \begin{bmatrix} A(\cdot) & B(\cdot) \\ C(\cdot) & D(\cdot) \end{bmatrix} \begin{bmatrix} X(\cdot) \\ U(\cdot) \end{bmatrix}. \tag{5.25}$$

Proof. The proof follows from Prop. 5.3 and Prop. 5.6. □

We now introduce the algorithm for LTV self-adjoint systems in Algorithm 5.

Algorithm 5 Duality-based algorithm for the self-adjoint case

Input: $(u_i(\cdot), y_i(\cdot))\ i = 1, .., N$
Output: $(A(\cdot), B(\cdot), C(\cdot), D(\cdot))$
Assumptions: Genericity, $n < N$

1. Build $E(\cdot)$, using (5.20):

$$e_{k,\ell}(t) := \int_0^t u_k(\tau)^\top y_\ell(\tau) + y_k(\tau)^\top u_\ell(\tau)\ d\tau \quad k, \ell = 1, .., N,$$
$$E(t) = \begin{bmatrix} e_{1,1}(t) & \cdot & e_{1,N} \\ \vdots & . & \vdots \\ e_{N,1}(t) & \cdot & e_{N,N} \end{bmatrix}.$$

2. Compute the state dimension: $\operatorname{rank}(E(\cdot)) = n$.

3. Partition $E(\cdot)$ as
$$E(\cdot) := \begin{bmatrix} E_{1,1}(\cdot) & E_{1,2}(\cdot) \\ E_{1,2}(\cdot)^\top & E_{2,2}(\cdot) \end{bmatrix}$$
such that $E_{1,1}(\cdot) \in \mathcal{A}^{n\times n}$ is generically nonsingular.

4. Define $X(\cdot) := [I_n \quad E_{1,1}(\cdot)^{-1}E_{1,2}(\cdot)]$, $Y(\cdot) := [y_1(\cdot)\ \cdots\ y_N(\cdot)]$ and $U(\cdot) := [u_1(\cdot)\ \cdots\ u_N(\cdot)]$.

5. Compute $\overset{\bullet}{X}$.

6. Compute $N(\cdot) \in \mathcal{A}^{N\times(n+m)}$ such that
$$\begin{bmatrix} X(\cdot) \\ U(\cdot) \end{bmatrix} N(\cdot) = I_{n+m}.$$

7. Define $(A(\cdot), B(\cdot), C(\cdot), D(\cdot))$ from (5.18).

5.5 Summary

In our approach we use *duality* to exploit the fact that external properties, i.e. inputs and outputs, are reflected in the internal ones, i.e. states. This allow us to compute the state trajectories from the external variables. It was proven that the state trajectories computed correspond to an i-s-o representation of the system to-be-identified. Then these state trajectories are used to compute the matrices $A(\cdot), B(\cdot), C(\cdot)$ and $D(\cdot)$ corresponding to such system. Sufficient conditions for the computation of such matrices are also given. Note that we do not impose conditions in the time variation properties.

In the first part, we treated the general case when the input-output trajectories of the dual and the primal are available. Then the matrix $E(\cdot)$ defined by (5.2) is factorized as in Prop. 5.5. Note that the choice of the basis of the state space is based on the factorization in 5.5.

We also provided a procedure for identification when the system is self-adjoint. In this case, we exploit the fact that the input-output trajectories of the primal are also trajectories of the dual.

Chapter 6

Identification of Hamiltonian Systems

In this chapter, we introduce a procedure for the identification of Hamiltonian systems using the duality-based approach and RKHS for sampled data developed in the previous chapters.

6.1 Hamiltonianity and self-adjointness

Self-adjoint systems arise when considering the *variational system* of a nonlinear Hamiltonian system. A Hamiltonian system is a dynamical system described by the scalar function $H(q, p, u)$, a function of canonical coordinates (q, p), and input u. H is called the Hamiltonian, and it often corresponds to the total energy of the system. Its time evolution is uniquely determined by Hamilton's equations:

$$\begin{aligned} \frac{d}{dt}q &= \frac{\partial H}{\partial p}(q, p, u) \\ \frac{d}{dt}p &= -\frac{\partial H}{\partial q}(q, p, u) \\ y &= \frac{\partial H}{\partial u}(q, p, u). \end{aligned} \tag{6.1}$$

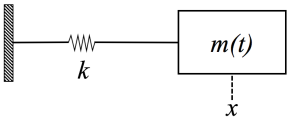


Figure 6.1: Harmonic oscillator.

An example of a Hamiltonian formulation in the LTV case is the harmonic oscillator with a time-varying mass. The harmonic oscillator is a very important system in physics, it occurs widely in nature. An example of this fact is that any mass subject to forces in a stable equilibrium acts as a harmonic oscillator for small vibrations.

Consider 1-D harmonic oscillator with an external force $u(\cdot)$, see Fig. 6.1. The total energy of the system is given by

$$H(q,p,u) = \frac{p^2}{2m} + \frac{kq^2}{2m} - uq$$

where $q = mx$, and $p = \frac{d}{dt}(mx)$ the momentum. The equations of motion are then given by (6.2),

$$\begin{aligned} \frac{d}{dt}q &= \frac{\partial H}{\partial p}(q,p,u) = p \\ \frac{d}{dt}p &= -\frac{\partial H}{\partial q}(q,p,u) = -\frac{kq}{m} + u \\ y &= -\frac{\partial H}{\partial u}(q,p,u) = q. \end{aligned} \tag{6.2}$$

where in the conventional form this is translated to

$$u = kx + \frac{d^2}{dt^2}(mx).$$

Now consider (6.1) in the nonlinear case. Let $x = \mathrm{col}(q,p)$, then (6.1) can be written as

$$\begin{aligned} \frac{d}{dt}x &= f(x,u) \\ y &= h(x) \end{aligned} \tag{6.3}$$

where (6.3) are state equations of a nonlinear system with $x \in \mathbb{R}^n$, $u, y \in \mathbb{R}^m$. Let $(\widetilde{u}(\cdot), \widetilde{x}(\cdot), \widetilde{y}(\cdot))$ be a solution of (6.3) and define

$$\begin{aligned} F(\cdot) &:= \frac{\partial f}{\partial x}(\widetilde{x}(\cdot), \widetilde{u}(\cdot)) \\ G(\cdot) &:= \frac{\partial f}{\partial u}(\widetilde{x}(\cdot), \widetilde{u}(\cdot)) \\ H(\cdot) &:= \frac{\partial h}{\partial x}(\widetilde{x}(\cdot)) \ . \end{aligned} \tag{6.4}$$

Now consider the *variational system* associated with (6.3) in a neighbourhood of the trajectory $(\widetilde{u}(\cdot), \widetilde{y}(\cdot), \widetilde{x}(\cdot))$, i.e.

$$\begin{aligned} \frac{d}{dt}x_v(\cdot) &= F(\cdot)x_v(\cdot) + G(\cdot)u_v(\cdot) \\ y_v(\cdot) &= H(\cdot)x_v(\cdot) \ , \end{aligned} \tag{6.5}$$

where $x_v \in \mathbb{R}^n$ and $y_v, u_v \in \mathbb{R}^m$ are the *variational state*, *output* and *input*, respectively.

The *adjoint system* of (6.5) is defined by

$$\begin{aligned} \frac{d}{dt}x_a(\cdot) &= -F(\cdot)^\top x_a(\cdot) - H(\cdot)^\top u_a(\cdot) \\ y_a(\cdot) &= G(\cdot)^\top x_a(\cdot)\,, \end{aligned} \tag{6.6}$$

with $x_a \in \mathbb{R}^n$, $y_a, u_a \in \mathbb{R}^m$ the *adjoint variational state*, *output* and *input*, respectively.

The variational system (6.5) is a linear approximation along a solution $(\widetilde{u}(\cdot), \widetilde{y}(\cdot), \widetilde{x}(\cdot))$ of the original nonlinear system (6.3) in the following sense. Consider the nonlinear system in (6.3) and the solution (u, y, x). Now let a small variation δu be given by $\delta u := u - u^*$, then the induced output and state of the system when a variation is applied are defined as $\delta y := y - y^*$ and $\delta x := x - x^*$, respectively. It follows that the nonlinear system with input u^* is given by

$$\begin{aligned} \frac{d}{dt}x^* &= f(x^*, u^*) \\ y^* &= h(x^*). \end{aligned}$$

This implies that

$$\frac{d}{dt}\delta x = f(x, u) - f(x^*, u^*). \tag{6.7}$$

Now using Taylor series for $f(x, u)$ in (6.7) and denoting higher order terms as O_f we have

$$\frac{d}{dt}\delta x = \left[f(x^*, u^*) + \frac{\partial f}{\partial x}(x^*, u^*)\delta x + \frac{\partial f}{\partial u}(x^*, u^*)\delta u + O_f\right] - f(x^*, u^*). \tag{6.8}$$

Since the variations are assumed to be "small", i.e. x is close enough to x^*, then the higher order terms in (6.8) are close to zero. Simplifying and neglecting higher order terms it follows that

$$\frac{d}{dt}\delta x(\cdot) \cong \frac{\partial f}{\partial x}(x^*(\cdot), u^*(\cdot))\delta x(\cdot) + \frac{\partial f}{\partial u}(x^*(\cdot), u^*(\cdot))\delta u(\cdot).$$

Following a similar procedure for the output equation we obtain

$$\begin{aligned} \delta y &= y - y^* \\ &= h(x) - h(x^*) \\ &= [h(x^*) + \frac{\partial h}{\partial x}(x^*)\delta x + O_h] - h(x*). \end{aligned}$$

Simplifying and neglecting the higher order terms O_h it follows that the output equation is now given by

$$\delta y(\cdot) \cong \frac{\partial h}{\partial x}(x^*)\delta x(\cdot).$$

Then the resulting variational system is given by

$$\begin{aligned}\frac{d}{dt}\delta x(\cdot) &= F(\cdot)\delta x(\cdot) + H(\cdot)\delta u(\cdot)\\ \delta y(\cdot) &= G(\cdot)\delta x(\cdot),\end{aligned}$$

with matrices $F(\cdot) := \frac{\partial f}{\partial x}(x^*(\cdot), u^*(\cdot))$, $H(\cdot) := \frac{\partial f}{\partial u}(x^*(\cdot), u^*(\cdot))$ and $G(\cdot) = \frac{\partial h}{\partial x}(x^*(\cdot))$.

It is shown in Crouch et al. (1995), pp. 608 proof of Theorem 2, that if the nonlinear system (6.3) is Hamiltonian, then every variational system along trajectories resulting from piecewise constant inputs is self-adjoint. This implies that the system in (6.5) has another internal representation as a linear Hamiltonian control system, and that the input-output map of the variational and its adjoint coincide.

For Hamiltonian systems Def. 4.1 coincides with Def. 1 in Crouch et al. (1995) with minus sign, i.e. for all trajectories $(x_v(\cdot), u_v(\cdot), y_v(\cdot))$ of the variational system and $(x_a(\cdot), u_a(\cdot), y_a(\cdot))$ of the adjoint system

$$u_v^\top y_a - y_v^\top u_a = \frac{d}{dt}\left(x_v^\top x_a\right) \tag{6.9}$$

holds.

Note also that the result of Prop. 4.4 now becomes

$$\int_{-\infty}^{+\infty} u_v(\tau)^\top y_a(\tau) - y_v(\tau)^\top u_a(\tau)\, d\tau = 0. \tag{6.10}$$

for all compact support trajectories (x_v, u_v, y_v) and (x_a, u_a, y_a).

From (6.9) together with the variational system representation in (6.5) and the adjoint variational system representation in (6.6) we can conclude that the representations are matched in the sense of Prop. 4.9 as follows.

Proposition 6.1. *Let (x_v, u_v, y_v) and (x_a, u_a, y_a) be trajectories satisfying the variational system representation in (6.5) and the adjoint variational system representation in (6.6) respectively. Then*

$$u_v^\top y_a - y_v^\top u_a = \frac{d}{dt}\left(x_v^\top x_a\right)$$

holds.

Proof. From (6.5) and (6.6) the claim follows from the following chain of equalities

$$\begin{aligned}
u_v^\top y_a - y_v^\top u_a &= u_v^\top (G^\top x_a) - (Hx_v)^\top u_a \\
&= (Gu_v)^\top x_a - x_v^\top (H^\top u_a) \\
&= (\frac{d}{dt}x_v - Fx_v)^\top x_a - x_v^\top (-\frac{d}{dt}x_a - F^\top x_a) \\
&= (\frac{d}{dt}x_v^\top - x_v^\top F^\top)x_a + x_v^\top (\frac{d}{dt}x_a + F^\top x_a) \\
&= (\frac{d}{dt}x_v^\top)x_a + x_v^\top (\frac{d}{dt}x_a) \\
&= \frac{d}{dt}(x_v^\top x_a).
\end{aligned}$$

□

Note that in the case when (6.5) corresponds to an i-s-o representation of a LTV system, the conditions of self-adjointness also hold in a straightforward manner, i.e. if a LTV system is Hamiltonian then it is also self-adjoint.

6.2 From input-output data to RKHS

In this section we develop an algorithm using RKHS to construct functions for sampled data. This will allow us to make the connection between sampled input-output data to continuous functions of time for performing the duality-based identification approach. In this work, we use the exponential kernel to embed the differentiability properties of the kernel into the functions.

6.2.1 The exponential kernel

As we have seen in section 3.2, the functions in a RKHS are defined by a linear combination of the kernel associated to it, see eq. 3.20. Thus, the properties of the functions in a RKHS are determined by the kernel. The exponential kernel, also called the Gaussian kernel, is widely used in RKHS theory due to its smoothness.

In our approach we consider the exponential kernel to exploit the properties differentiability and closeness to differentiation. These will allow us to define analytic functions needed for our approach and to improve the computational efficiency.

Consider the exponential function

$$k(x, x') := \exp\left(-\frac{\|x - x'\|^2}{\mu^2}\right)$$

for some $\mu \in \mathbb{R}$ and $x, x' \in X$.

As shown in Def. 3.9, every kernel is formed by an inner product between feature maps. Then, to prove that the exponential function is a kernel, and consequently positive definite, we need to demonstrate that the function is formed by the inner product of feature maps.

Proposition 6.2. *Let X be a normed vector space, then for all $x, x' \in X$ and $\mu \in \mathbb{R}$ the function defined by*

$$k(x, x') := \exp\left(-\frac{\|x - x'\|^2}{\mu^2}\right) \tag{6.11}$$

is a kernel.

Proof. Consider $\langle \cdot, \cdot \rangle : X \times X \to \mathbb{R}$. It is easy to see that $\langle x, x' \rangle$ is a kernel since $\langle x, x \rangle \geq 0$ and it is symmetric. Moreover, from Prop. 3.14, $\exp\left(c\langle x, x' \rangle\right)$ is also a kernel with $c \in \mathbb{R}^+$, then there exist $\Phi_1 : X \to H$ s.t.

$$k_1(x, x') := \langle \Phi_1(x), \Phi_1(x') \rangle = \exp\left(c\langle x, x' \rangle\right). \tag{6.12}$$

According to Shawe-Taylor and Cristianini (2004) (see p. 40 section 2.3.2 of this reference), we can create new kernels by normalizing them. Let

$$k_{norm}(x, x') := \left\langle \frac{\Phi_1(x)}{\|\Phi_1(x)\|}, \frac{\Phi_1(x')}{\|\Phi_1(x')\|} \right\rangle = \frac{k_1(x, x')}{\sqrt{k_1(x, x) k_1(x', x')}}. \tag{6.13}$$

Then the normalized kernel is defined by

$$\begin{aligned} k_{norm}(x, x') &= \frac{\exp\left(c\langle x, x' \rangle\right)}{\sqrt{\exp\left(c\langle x, x \rangle\right) \exp\left(c\langle x', x' \rangle\right)}} \\ &= \exp\left(c\langle x, x' \rangle - \frac{c\langle x, x \rangle}{2} - \frac{c\langle x', x' \rangle}{2}\right). \end{aligned} \tag{6.14}$$

Now define $c := 2/\mu^2$, then

$$k_{norm}(x, x') = \exp\left(\frac{2\langle x, x' \rangle}{\mu^2} - \frac{\langle x, x \rangle}{\mu^2} - \frac{\langle x', x' \rangle}{\mu^2}\right). \tag{6.15}$$

Since

$$\langle x - x', x - x' \rangle = \langle x, x \rangle - 2\langle x, x' \rangle + \langle x', x' \rangle.$$

and $\|x\| = \sqrt{\langle x, x \rangle}$, then

$$k_{norm}(x, x') = \exp\left(-\frac{\|x - x'\|^2}{\mu^2}\right) \tag{6.16}$$

is a kernel. Moreover it is an inner product between normalized feature maps, and consequently it is positive definite. □

It follows from Th. 3.18 that the exponential kernel in (6.11) is a reproducing kernel defining a unique RKHS given by

$$\mathcal{H} := \left\{ f(\cdot) \mid f(\cdot) = \sum_{i=1}^{l} \alpha_i \exp\left(-\frac{\| \cdot - x_i \|^2}{\mu^2} \right) \right\} \tag{6.17}$$

equipped with the inner product

$$\langle f, g \rangle_{\mathcal{H}} = \sum_{i=1}^{l} \sum_{j=1}^{l} \alpha_i \beta_j k(x_i, x_j). \tag{6.18}$$

for some $\mu \in \mathbb{R}$ and given $x_i \in X$ $i = 1, .., l$.

It follows from Th. (3.19) that the differentiability properties of every function in the RKHS are given by the differentiability properties of the associated kernel.

To show infinitely differentiability in RKHS associated with a specific kernel we just need to verify the differentiability of the kernel. This is stated in the following theorem.

Theorem 6.3. *Let k be an infinitely differentiable reproducing kernel, then its corresponding reproducing kernel Hilbert space is infinitely differentiable. The converse is also true.*

Proof. From Def. 3.16 if $\mathcal{H}$ is a RKHS, then a reproducing kernel k in $\mathcal{H}$ must exist. It follows from Th. 3.19 that if k is infinite differentiable then every function in $\mathcal{H}$ is infinite differentiable.

To prove the second part of the claim, let the RKHS $\mathcal{H}$ be infinite differentiable, and assume that the kernel k in $\mathcal{H}$ has derivatives up to order n. From Cor. 3.19 it follows that every function in $\mathcal{H}$ is n-times continuously differentiable, then the corresponding RKHS is n-times continuously differentiable which is a contradiction. This concludes the proof. □

Now we define the differentiability properties for the RKHS associated to the exponential kernel in the following proposition.

Proposition 6.4. *Let $X \subseteq \mathbb{R}$, and*

$$K(x, x') = e^{-\left(\frac{\|x - x'\|^2}{\mu^2} \right)} \quad \forall x, x' \in X, \quad \mu \in \mathbb{R}$$

be the reproducing kernel defining the reproducing kernel Hilbert space

$$\mathcal{H} := \left\{ f(\cdot) \mid f(\cdot) = \sum_{i=1}^{l} \alpha_i e^{-\left(\frac{\|\cdot - x_i\|^2}{\mu^2} \right)} \right\}$$

for $x_1, .., x_l \in X$ and $\alpha_1, .., \alpha_l \in \mathbb{R}$. Then every function in the reproducing kernel Hilbert space is infinitely differentiable.

Proof. Since every function in $\mathcal{H}$ is defined by

$$f(\cdot) = \sum_{i=1}^{l} \alpha_i e^{-\left(\frac{\|\cdot - x_i\|^2}{\mu^2}\right)},$$

we can prove the differentiability of every function from the following chain of equalities:

$$\begin{aligned}
f(x) &= \sum_{i=1}^{l} \alpha_i e^{-\left(\frac{\|x - x_i\|^2}{\mu^2}\right)} \\
f(x)^{(1)} &= \sum_{i=1}^{l} \alpha_i e^{-\left(\frac{\|x - x_i\|^2}{\mu^2}\right)} \left(-\frac{2(x - x_i)}{\mu^2}\right) \\
f(x)^{(2)} &= \sum_{i=1}^{l} \alpha_i e^{-\left(\frac{\|x - x_i\|^2}{\mu^2}\right)} \left[\left(-\frac{2}{\mu^2}\right) + \left(-\frac{2(x - x_i)}{\mu^2}\right)^2\right] \\
&\vdots \\
f(x)^{(n)} &= \sum_{i=1}^{l} \alpha_i e^{-\left(\frac{\|x - x_i\|^2}{\mu^2}\right)} P(x, x_i, \mu),
\end{aligned}$$

where $P(x, x_i, \mu)$ is a polynomial of degree n. Then $f(\cdot)$ is infinitely differentiable since $f(\cdot)^{(n)}$ for an arbitrary n always exists. Therefore every function in the RKHS is infinitely differentiable. This concludes the proof. □

6.2.2 Infinite differentiable RKHS

Consider the following set of given data

$$\{(\mathbf{u}_j, t_j)_{j=1,..,T} \mid \mathbf{u}_j, t_j \in \mathbb{R} \ \text{ for } j = 1, .., T \ j, T \in \mathbb{N}\} \tag{6.19}$$

where $\mathbf{u}_j$ is the $j-$th sample at the $j-$th time t for $j = 1, .., T$. From (6.19) we want to define a continuous function of time $u(\cdot)$ for $\mathbf{u}_j$ such that $u(t_j) = \mathbf{u}_j$ using RKHS.

Let the exponential kernel in (6.11) and the set in (6.19). It follows that, for some $\mu \in \mathbb{R}$, the RKHS associated to the exponential kernel is given by

$$\mathcal{H} := \left\{ f(\cdot) \mid f(\cdot) = \sum_{i=1}^{N} \alpha_i \exp\left(-\frac{\|\cdot - t_i\|^2}{\mu^2}\right) \right\}$$

equipped with the inner product

$$\langle f, g \rangle_{\mathcal{H}} = \sum_{i=1}^{N} \sum_{j=1}^{N} \alpha_i \beta_j k(t_i, t_j).$$

It follows from eq. (3.25) that we can compute the coefficients $\alpha' s$ of f for the set (6.19) by solving a set of linear equations as in (2.24), e.g. for the entry $\mathbf{u}_j$ we solve

$$\begin{bmatrix} \mathbf{u}_1 \\ \mathbf{u}_2 \\ \vdots \\ \mathbf{u}_T \end{bmatrix} = \begin{bmatrix} k(t_1, t_1) & k(t_1, t_2) & \cdots & k(t_1, t_T) \\ k(t_2, t_1) & k(t_2, t_2) & \cdots & k(t_2, t_T) \\ \vdots & \vdots & \ddots & \vdots \\ k(t_T, t_1) & k(t_T, t_2) & \cdots & k(t_T, t_T) \end{bmatrix} \begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_T \end{bmatrix} \tag{6.20}$$

for $\text{col}(\alpha_1, .., \alpha_T)$.

Then we can define the functions for the data set (6.19) as

$$u(\cdot) := \sum_{j=1}^{T} \alpha_j \exp\left(-\frac{\| \cdot - t_j \|^2}{\mu^2} \right).$$

Remark 6.5. In the Gaussian kernel, the parameter μ corresponds to the width of the Gaussian function, i.e. it defines the average Euclidean distance that measures the dispersion of the data. In the case of time functions with uniform sample time, the parameter corresponds to the sample time of the data, i.e. $\mu = t_{i+1} - t_i$.

6.3 Duality-based identification with RKHS

In Chapter 5 section 5.4, we developed a procedure to identify self-adjoint systems. From Crouch et al. (1995), we know that if a system is Hamiltonian then it has the self-adjointness property. Consequently every input-output trajectory of the primal is also a trajectory of the dual. In this section we introduce an identification procedure for Hamiltonian systems using the theory developed previously, i.e. the duality approach and RKHS.

First consider $\mathbf{u}_{i,j}$ be the $j-$th sample of the $\mathbf{u}$ data at time t_j for the $i-$th trajectory. In this context, the identification problem for Hamiltonian systems is stated as follows.

Identification problem. The Hamiltonian case.

Given

- a set of samples of N input-output trajectories

$$\{(\mathbf{u}_{i,j}, \mathbf{y}_{i,j}) \mid \mathbf{u}_{i,j}, \mathbf{y}_{i,j} \in \mathbb{R}^m \text{ for } j = 1, .., T \;\; j, T \in \mathbb{N}\}_{i=1,..,N},$$

generated by a Hamiltonian system.

Compute

- $(u_i(\cdot), y_i(\cdot))$ such that $u_i(t_j) := \mathbf{u}_{i,j}$ and $y_i(t_j) := \mathbf{y}_{i,j}$,
- matrices $A(\cdot), B(\cdot), C(\cdot), D(\cdot)$ corresponding to an unfalsified model for $(u_i(\cdot), y_i(\cdot))_{i=1,..,N}$

Since the LTV system is self-adjoint we use (4.8) for our procedure.

In the first stage of the algorithm we proceeded to construct the functions for the input-output data using RKHS. For $i = 1, .., N$ and $j = 1, .., T$. We then define

$$\begin{aligned} u_i(\cdot) &:= \sum_{j=1}^{N_2} \alpha_{i,j} \exp\left(-\frac{\|\cdot - t_j\|^2}{\mu^2}\right) \\ y_i(\cdot) &:= \sum_{j=1}^{N_2} \beta_{i,j} \exp\left(-\frac{\|\cdot - t_j\|^2}{\mu^2}\right) \end{aligned} \tag{6.21}$$

where $\alpha_{i,j}$ and $\beta_{i,j}$ are computed as in (6.20). Since we now defined $u_i(\cdot)$ and $y_i(\cdot)$, we can compute the entries of the energy matrix $E(\cdot)$ as in (5.20) as

$$\mathbf{e}_{k,\ell}(t) := \int_0^t u_k(\tau)^\top y_\ell(\tau) + y_k(\tau)^\top u_\ell(\tau)\, d\tau \quad k, \ell = 1, .., N. \tag{6.22}$$

Then the matrix $E(\cdot)$ is defined by

$$E(\cdot) := \begin{bmatrix} \mathbf{e}_{1,1}(\cdot) & \cdots & \mathbf{e}_{1,N}(\cdot) \\ \vdots & \ddots & \vdots \\ \mathbf{e}_{N,1}(\cdot) & \cdots & \mathbf{e}_{N,N}(\cdot) \end{bmatrix}.$$

From Prop. 5.3, we have that generically the rank of the matrix $E(\cdot)$ whose entries are defined by (5.20) equals the dimension of the state, i.e. $\operatorname{rank}(E(\cdot)) = n$.

Now consider the partition

$$E(\cdot) = \begin{bmatrix} \mathbf{E}_{1,1}(\cdot) & \mathbf{E}_{1,2}(\cdot) \\ \mathbf{E}_{1,2}(\cdot)^\top & \mathbf{E}_{2,2}(\cdot) \end{bmatrix} \tag{6.23}$$

where $\mathbf{E}_{1,1}(\cdot) \in \mathcal{H}^{n\times n}$, $\mathbf{E}_{1,2}(\cdot) \in \mathcal{H}^{n\times(N-n)}$, and $\mathbf{E}_{2,2}(\cdot) \in \mathcal{H}^{(N-n)\times(N-n)}$. It follows from 5.4 that $\mathbf{E}_{1,1}(\cdot)$ in (6.23) is nonsingular. Since $\operatorname{rank}(E(\cdot)) = n < N$, it follows that the factorization is given by Prop. 5.5, and consequently state trajectories can be defined as in Prop. 5.14, then

$$\begin{aligned} X(\cdot) &:= [I_n \ \mathbf{E}_{1,1}(\cdot)^{-\top}\mathbf{E}_{2,1}(\cdot)^\top], \\ Y(\cdot) &:= [y_1(\cdot) \ \cdots \ y_N(\cdot)], \\ U(\cdot) &:= [u_1(\cdot) \ \cdots \ u_N(\cdot)], \end{aligned}$$

where $\mathbf{E}_{1,1}(\cdot)$ is the n by n $(1,1)$non-singular block sub-matrix of $E(\cdot)$ and $\mathbf{E}_{2,1}(\cdot)$ is the n by $N-n$ $(2,1)$ block sub-matrix of $E(\cdot)$.

If sufficient conditions of Prop. 5.3, we compute a right inverse matrix $R(\cdot)$ such that

$$\begin{bmatrix} X(\cdot) \\ U(\cdot) \end{bmatrix} R(\cdot) = I_{n+m}.$$

Then we solve

$$\begin{bmatrix} \dot{X}(\cdot) \\ Y(\cdot) \end{bmatrix} R(\cdot) = \begin{bmatrix} A(\cdot) & B(\cdot) \\ C(\cdot) & D(\cdot) \end{bmatrix}.$$

for $(A(\cdot), B(\cdot), C(\cdot), D(\cdot))$.

Then the algorithm for duality-based identification using RKHS is summarized in Algorithm 6.

Algorithm 6 Duality-based algorithm using RKHS

Input: $(\mathbf{u}_{i,j}, \mathbf{y}_{i,j})$ for $j = 1, .., T$ and $i = 1, .., N$

Output: $(A(\cdot), B(\cdot), C(\cdot), D(\cdot))$

Assumptions: Genericity, Prop. 5.3, $T > 1$, $N > n$ where n is the state dimension

1. Define $(u_i(\cdot), y_i(\cdot))$ using RKHS as in (6.21).
2. Compute the energy matrix $E(\cdot)$ as in (6.22).
3. Define $E(\cdot) := [\mathbf{e}_{k,\ell}(\cdot)]_{k,\ell=1,..,N}$.
4. Compute the state dimension: $\operatorname{rank}(E(\cdot)) = n$.
5. Partition $E(\cdot)$ as
$$E(\cdot) := \begin{bmatrix} \mathbf{E}_{1,1}(\cdot) & \mathbf{E}_{1,2}(\cdot) \\ \mathbf{E}_{2,1}(\cdot) & \mathbf{E}_{2,2}(\cdot) \end{bmatrix}$$
such that $\mathbf{E}_{1,1}(\cdot) \in \mathbb{R}^{n \times n}$.
6. Define
$$\begin{aligned} X(\cdot) &:= [I_n \;\; \mathbf{E}_{1,1}(\cdot)^{-\top} \mathbf{E}_{2,1}(\cdot)^{\top}], \\ Y(\cdot) &:= [y_1(\cdot) \;\; \cdots \;\; y_N(\cdot)], \\ U(\cdot) &:= [u_1(\cdot) \;\; \cdots \;\; u_N(\cdot)]. \end{aligned}$$
7. Compute $\overset{\bullet}{X}$.
8. Compute $R(\cdot)$ such that
$$\begin{bmatrix} X(\cdot) \\ U(\cdot) \end{bmatrix} R(\cdot) = I_{n+m}.$$
9. Compute $(A(\cdot), B(\cdot), C(\cdot), D(\cdot))$ from
$$\begin{bmatrix} \overset{\bullet}{X}(\cdot) \\ Y(\cdot) \end{bmatrix} R(\cdot) = \begin{bmatrix} A(\cdot) & B(\cdot) \\ C(\cdot) & D(\cdot) \end{bmatrix}.$$

Chapter 7

Implementation and Validation

In this chapter we implement the theory and algorithms developed in this thesis. We use real data to identify a state space representation of a physical LTV self-adjoint system. First, RKHS is used for building continuous functions for the input-output trajectories. Then, the duality-based algorithm is applied.

7.1 The electrical oscillator case

Consider the LTV system (electrical oscillator) in Fig. 7.1 with a time-varying inductor L and a constant capacitor C. In this example the input variable $u(\cdot)$ is the input voltage and the output $q(\cdot)$ is the charge of the capacitor.

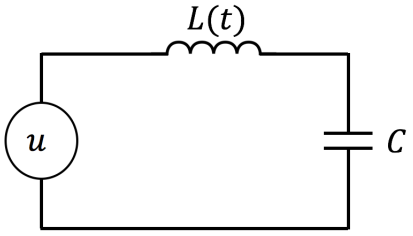


Figure 7.1: LTV LC system

The input-output equation of an LC system with time varying inductor is given by

$$\left(\frac{d}{dt}L\right)\left(\frac{d}{dt}q\right) + L\left(\frac{d^2}{dt^2}q\right) + \frac{1}{C}q - u = 0. \tag{7.1}$$

In this example, to verify that the system in (7.1) is Hamiltonian and consequently self-adjoint, we use the conditions of Hamiltonianity introduced in Crouch and Van der Schaft (1987).

For systems described by

$$F\left(y, \frac{d}{dt}y, \frac{d^2}{dt^2}y, u, \frac{d}{dt}u\right) = 0 \in \mathbb{R}, \tag{7.2}$$

with $y, u \in \mathbb{R}$, the conditions for Hamiltonianity were studied in Crouch and Van der Schaft (1987). They showed that the system (7.2) is Hamiltonian if

1. $\frac{\partial F}{\partial(\frac{d}{dt}u)} = 0$
2. $\frac{\partial F}{\partial u}\left(\frac{d}{dt}\frac{\partial F}{\partial(\frac{d^2}{dt^2}y)} - \frac{\partial F}{\partial(\frac{d}{dt}y)}\right) = \frac{\partial F}{\partial(\frac{d^2}{dt^2}y)}\frac{d}{dt}\frac{\partial F}{\partial u}$

hold for every solution satisfying (7.2).

For (7.1), we have that the condition 1 of Hamiltonianity holds since there is no term containing $\frac{d}{dt}u$. Then condition 2 becomes

$$\frac{\partial F}{\partial u}\left(\frac{d}{dt}\frac{\partial F}{\partial(\frac{d^2}{dt^2}y)} - \frac{\partial F}{\partial(\frac{d}{dt}y)}\right) = -1\left(\frac{d}{dt}L - \frac{d}{dt}L\right) = 0$$

which shows that the system (7.1) is Hamiltonian.

The system (7.1) in state space representation can be written as

$$\begin{aligned} \frac{d}{dt}\begin{bmatrix} q(\cdot) \\ \frac{d}{dt}q(\cdot) \end{bmatrix} &= \begin{bmatrix} 0 & 1 \\ -\frac{1}{L(\cdot)C} & -\frac{1}{L(\cdot)}\frac{d}{dt}L(\cdot) \end{bmatrix} + \begin{bmatrix} 0 \\ \frac{1}{L(\cdot)} \end{bmatrix} u(\cdot) \\ y(\cdot) &= \begin{bmatrix} 1 & 0 \end{bmatrix}\begin{bmatrix} q \\ \frac{d}{dt}q \end{bmatrix} \end{aligned} \tag{7.3}$$

7.2 Simulation

Let the circuit in Figure 7.1 be the to be identified LTV system. The input-output variables are u, the input voltage, and q, the charge of the capacitor, respectively. To generate the input-output trajectories the circuit in Figure 7.1 is simulated in MATLAB SIMULINK with a sample rate of $1e-4$ *sec* and a simulation time of 1 *sec*. Note that the sample rate was chosen such that it captures enough information of the dynamics of the signals, this criteria will vary according to the types of signals of each application. SIMSCAPE TOOLBOX is used to accurately implement the electronic components. The data is managed in a discrete form to simulate a real scenario. The data is stored in MATLAB time series objects to simplify the computational complexity.

Figure 7.2 shows the set-up of the simulated experiment. The *TV L signal* block induces a time-varying inductance signal to the inductor $L(t)$. The time-varying inductance signal is simulated with the function $L(t) = 0.001 * sin(2\pi 3t) + 0.003\ H$ and the capacitance

with $C = 1000\ \mu F$. The *Input* block produces four different inputs which are going to generate four different experiments. The results are stored as MATLAB time series objects using the *Input storage* block and the *output storage* block. The *Input storage* block stores the input voltage and the *output storage* block stores the charge in the capacitor that is measured indirectly through the voltage in the capacitor (i.e. $q = CV_c$). The *SPS* blocks are used to convert the unitless SIMULINK input signal to a simulated physical signal.

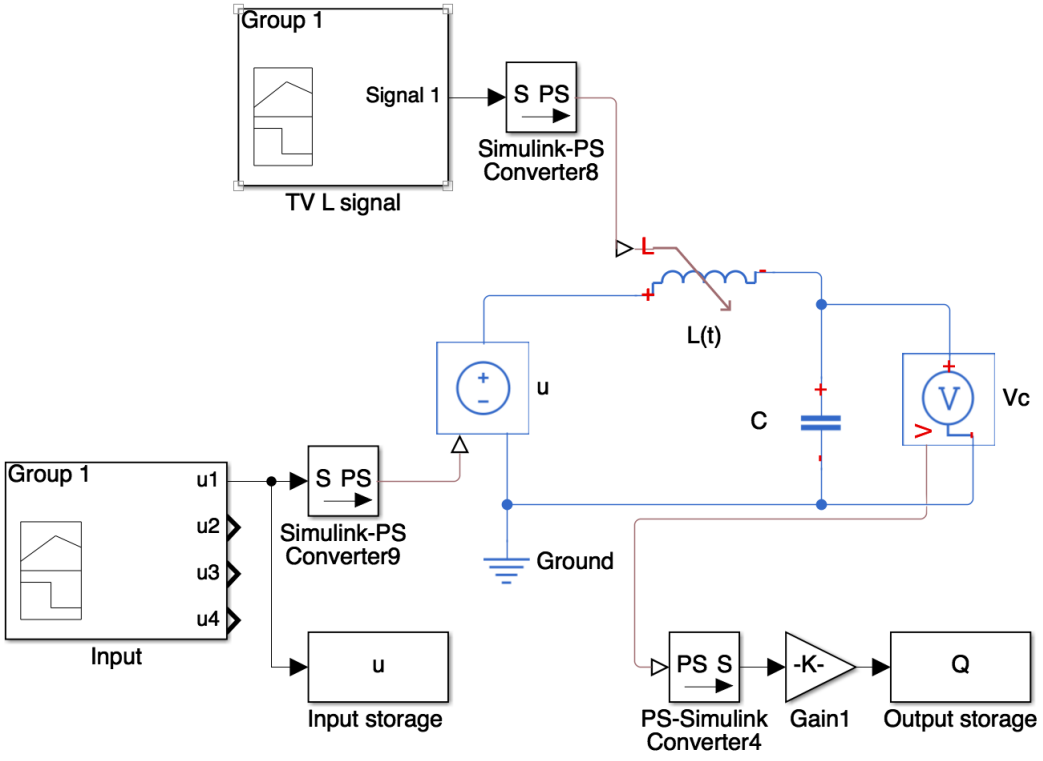


Figure 7.2: SIMULINK simulated experiment set-up for the LTV electrical oscillator.

Four different experiments using four inputs are generated in this simulation and stored as $(u_i(\cdot), y_i(\cdot))_{i=1,..,4}$. The inputs are linear combinations of sinusoidal signals chosen randomly to excite the dynamics of the system.

We then use these input-output trajectories to identify an unfalsified model for the data using the duality-based approach.

7.2.1 Identification

For the identification procedure we will use three input-output trajectories, i.e. $N = 3$, and the forth input-output trajectory will be used for validation.

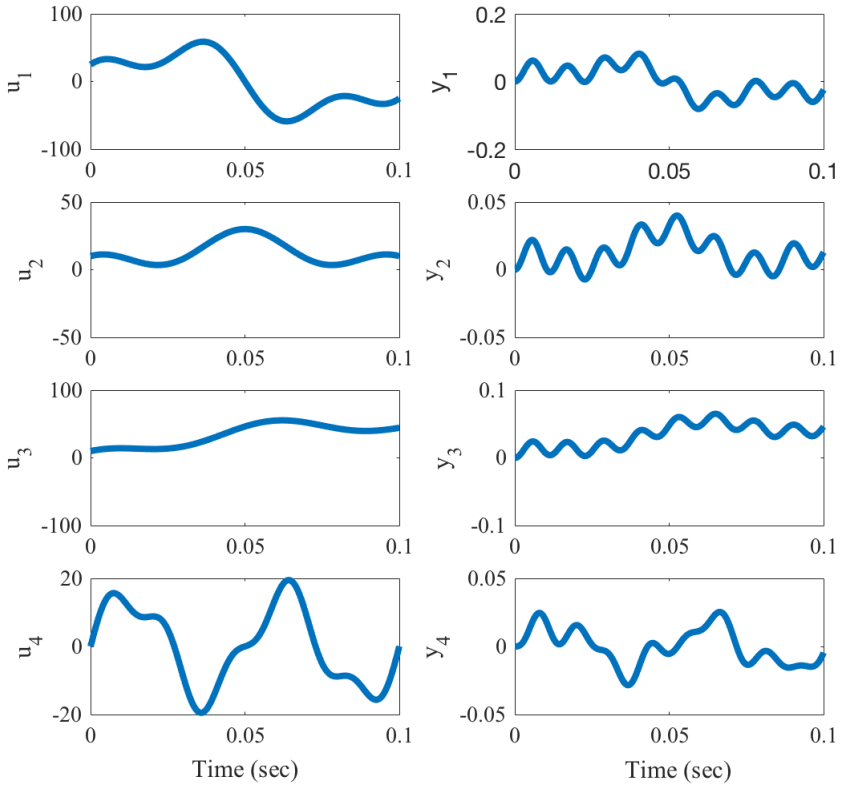


Figure 7.3: Input-output trajectories.

RKHS for data

In the first stage of the algorithm we construct the functions for the input-output data using RKHS based on Algorithm 6. For this, we choose the exponential kernel described in Section 6.2.1 exploiting the differentiability properties and closeness under differentiation.

Consider 3 input-output trajectories, i.e. $i = 1, 2, 3$, and let $T = 2000$ be the number of samples for each trajectory. We now define

$$\widehat{u}_i(\cdot) := \sum_{j=1}^{T} \alpha_{i,j} \exp\left(-\frac{\|\cdot - t_j\|^2}{\mu^2}\right)$$

$$\widehat{y}_i(\cdot) := \sum_{j=1}^{T} \beta_{i,j} \exp\left(-\frac{\|\cdot - t_j\|^2}{\mu^2}\right)$$

where t_j is the sampled time with t_1 the initial sample, and $\alpha_{i,j}$ and $\beta_{i,j}$ are constant coefficients computed as in (6.20), i.e. we solve

$$\begin{bmatrix} u_i(t_1) \\ u_i(t_2) \\ \vdots \\ u_i(t_T) \end{bmatrix} = \begin{bmatrix} k(t_1,t_1) & k(t_1,t_2) & \cdots & k(t_1,t_T) \\ k(t_2,t_1) & k(t_2,t_2) & \cdots & k(t_2,t_T) \\ \vdots & \vdots & \ddots & \vdots \\ k(t_T,t_1) & k(t_T,t_2) & \cdots & k(t_T,t_T) \end{bmatrix} \begin{bmatrix} \alpha_{i,1} \\ \alpha_{i,2} \\ \vdots \\ \alpha_{i,T} \end{bmatrix} \tag{7.4}$$

for $\text{col}(\alpha_{i,1}, .., \alpha_{i,T})$. The same procedure is followed for the output equations.

The energy matrix

Since the LTV system is Hamiltonian we use (4.8) for the construction of the matrix $\mathbf{E}(\cdot)$ whose elements are given by

$$\mathbf{e}_{k,\ell}(t) := \int_0^t u_k(\tau)^\top y_\ell(\tau) - y_k(\tau)^\top u_\ell(\tau)\, d\tau \quad k,\ell = 1, .., 3. \tag{7.5}$$

We then define $\mathbf{E}(\cdot)$ as

$$\mathbf{E}(\cdot) := \begin{bmatrix} \mathbf{e}_{1,1}(\cdot) & \mathbf{e}_{1,2}(\cdot) & \mathbf{e}_{1,3}(\cdot) \\ \mathbf{e}_{2,1}(\cdot) & \mathbf{e}_{2,2}(\cdot) & \mathbf{e}_{2,3}(\cdot) \\ \mathbf{e}_{3,1}(\cdot) & \mathbf{e}_{3,2}(\cdot) & \mathbf{e}_{3,3}(\cdot) \end{bmatrix},$$

It follows from Prop. 5.3 that generically the rank of the matrix $\mathbf{E}(\cdot)$ whose entries are defined by (7.5) equals the dimension of the state, i.e. $\text{rank}(\mathbf{E}(\cdot)) = 2$. Then matrix $\mathbf{E}(\cdot)$ can be partitioned as

$$\mathbf{E}_{1,1}(\cdot) := \begin{bmatrix} \mathbf{e}_{1,1}(\cdot) & \mathbf{e}_{1,2}(\cdot) \\ \mathbf{e}_{2,1}(\cdot) & \mathbf{e}_{2,2}(\cdot) \end{bmatrix} \quad \mathbf{E}_{1,2}(\cdot) := \begin{bmatrix} \mathbf{e}_{1,3}(\cdot) \\ \mathbf{e}_{2,3}(\cdot) \end{bmatrix}$$

$$\mathbf{E}_{2,1}(\cdot) := \begin{bmatrix} \mathbf{e}_{3,1}(\cdot) & \mathbf{e}_{3,2}(\cdot) \end{bmatrix} \quad \mathbf{E}_{2,2}(\cdot) := \begin{bmatrix} \mathbf{e}_{3,3}(\cdot) \end{bmatrix}$$

System matrix computation

Since $\text{rank}(\mathbf{E}(\cdot)) = 2 < N$, it follows that the factorization is given by Prop. 5.5, and consequently state trajectories can be defined as in Prop. 5.14. Then defining

$$\begin{aligned} X(\cdot) &:= [I_n \;\; \mathbf{E}_{1,1}(\cdot)^{-\top}\mathbf{E}_{2,1}(\cdot)^\top], \\ Y(\cdot) &:= [\widehat{y}_1(\cdot)\; \widehat{y}_2(\cdot)\; \widehat{y}_3(\cdot)], \\ U(\cdot) &:= [\widehat{u}_1(\cdot)\; \widehat{u}_2(\cdot)\; \widehat{u}_3(\cdot)], \end{aligned}$$

it follows from Prop. 5.3 that we can computed $N(\cdot)$ such that

$$\begin{bmatrix} X(\cdot) \\ U(\cdot) \end{bmatrix} N(\cdot) = I_3.$$

Then we solve

$$\begin{bmatrix} \dot{X}(\cdot) \\ Y(\cdot) \end{bmatrix} N(\cdot) = \begin{bmatrix} A(\cdot) & B(\cdot) \\ C(\cdot) & D(\cdot) \end{bmatrix}.$$

for $(A(\cdot), B(\cdot), C(\cdot), D(\cdot))$.

Then, we obtain matrices $(A(\cdot), B(\cdot), C(\cdot), D(\cdot))$ of the unfalsified model for the given data.

Remark 7.1. Note that the matrices computed from the algorithm are not the same as the ones in (7.3). The reason is that the factorization of the state trajectories determines the choice of the basis of the state space, thus the basis may not be the same as the ones in (7.3).

7.2.2 Validation

For validation, we use an input-output trajectory not used for identification, i.e. in this case $(u_4(\cdot), y_4(\cdot))$ in Figure 7.3. We first solve the identified state space model with numerical integration in MATLAB. Then, we verify that the output trajectory of the identified model and the one simulated coincide. The results of the simulation and the identified model are shown in Fig. 7.4.

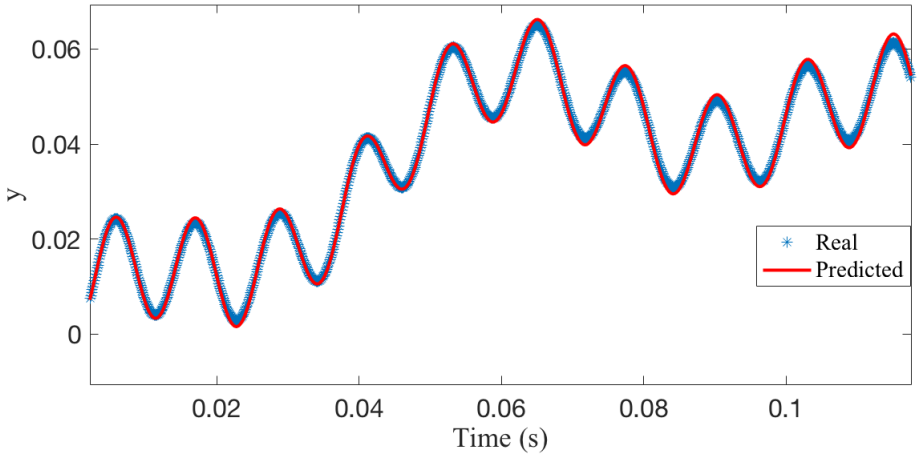


Figure 7.4: Output of the identified LTV system (red) and output of the real LTV system (blue).

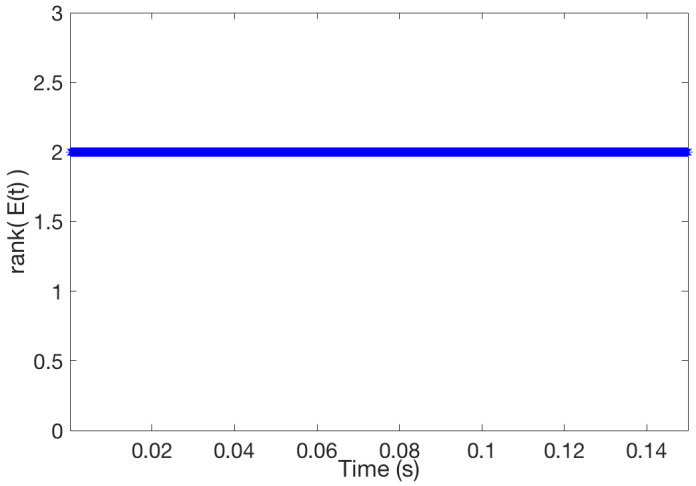


Figure 7.5: Rank of matrix $E(\cdot)$.

We also verify the rank of the matrix $E(\cdot)$ over the time interval to show the generic property. In this case, $\text{rank}(E(\cdot)) = 2$ which corresponds to the state dimension of the system (7.1), see Fig. 7.5.

7.3 Experimental setup

An experimental procedure is carried out to demonstrate our approach on a real-life system. The circuit used for the implementation is the one shown in Figure 7.6.

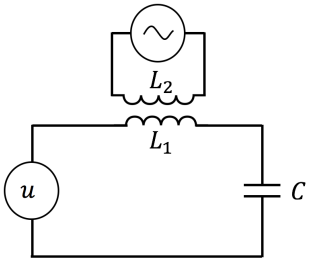


Figure 7.6: Diagram of the to-be-identified system.

A sinusoidal signal is applied to the inductor L_2. The inductor L_2 produces a magnetic field which influences the behaviour of L_1. This effect allows us to emulate a time-varying process in the inductor L_1.

For the experimental procedure the values of the parameters are: $L_1 = 10mH$, $L_2 = 10mH$ and $C = 680nF$. The input-output variables are u, the input voltage, and q, the charge in the capacitor, respectively.

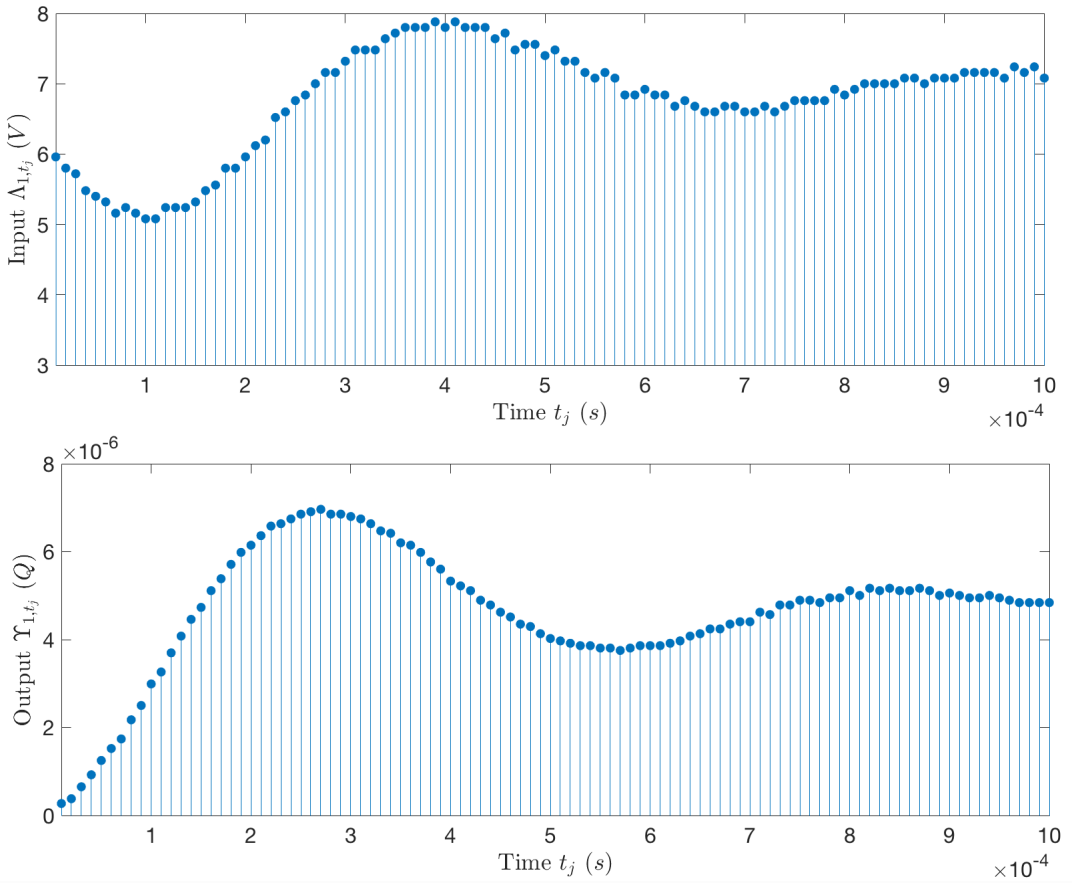


Figure 7.7: Sampled input Λ_{1,t_j} and output Υ_{1,t_j}.

We generated three outputs corresponding to three different inputs in physical experiments. 351 samples of these signals were stored as Λ_{i,t_j} for the input and Υ_{i,t_j} for the output of the i-th experiment with a sample rate of $10\mu s$, see e.g. Fig. 7.7 for Λ_{1,t_j}, and Υ_{1,t_j}.

7.3.1 Implementation

Algorithm 6 was implemented in MATLAB. In the first stage of the algorithm we proceeded to construct the functions for the input-output data using RKHS. For $i = 1, 2, 3$ and $T = 350$, we defined

$$\widehat{u}_i(\cdot) := \sum_{j=1}^{T} \alpha_{i,j} \exp\left(-\frac{\|\cdot - t_j\|^2}{\mu^2}\right)$$
$$\widehat{y}_i(\cdot) := \sum_{j=1}^{T} \beta_{i,j} \exp\left(-\frac{\|\cdot - t_j\|^2}{\mu^2}\right)$$

where $\alpha_{i,j}$ and $\beta_{i,j}$ are computed as in (6.20). We defined $\mathbf{E}(\cdot)$ as in step 3:

$$\mathbf{E}(\cdot) := \begin{bmatrix} \mathbf{e}_{1,1}(\cdot) & \mathbf{e}_{1,2}(\cdot) & \mathbf{e}_{1,3}(\cdot) \\ \mathbf{e}_{2,1}(\cdot) & \mathbf{e}_{2,2}(\cdot) & \mathbf{e}_{2,3}(\cdot) \\ \mathbf{e}_{3,1}(\cdot) & \mathbf{e}_{3,2}(\cdot) & \mathbf{e}_{3,3}(\cdot) \end{bmatrix},$$

and partitioned the matrix as follows:

$$\mathbf{E}_{1,1}(\cdot) := \begin{bmatrix} \mathbf{e}_{1,1}(\cdot) & \mathbf{e}_{1,2}(\cdot) \\ \mathbf{e}_{2,1}(\cdot) & \mathbf{e}_{2,2}(\cdot) \end{bmatrix} \quad \mathbf{E}_{1,2}(\cdot) := \begin{bmatrix} \mathbf{e}_{1,3}(\cdot) \\ \mathbf{e}_{2,3}(\cdot) \end{bmatrix}.$$

$$\mathbf{E}_{2,1}(\cdot) := \begin{bmatrix} \mathbf{e}_{3,1}(\cdot) & \mathbf{e}_{3,2}(\cdot) \end{bmatrix} \quad \mathbf{E}_{,3}(\cdot) := \begin{bmatrix} \mathbf{e}_{3,3}(\cdot) \end{bmatrix}$$

We computed the rank of $\mathbf{E}(\cdot)$, which in this case is $\operatorname{rank}(\mathbf{E}(\cdot)) = 2$, the dimension of the state. Since $\operatorname{rank}(\mathbf{E}(\cdot)) = n < N$, it follows that the factorization is given by Prop. 5.5, and consequently state trajectories can be defined as in Prop. 5.14. Then defining

$$X(\cdot) := [I_n \quad \mathbf{E}_{1,1}(\cdot)^{-\top} \mathbf{E}_{2,1}(\cdot)^{\top}],$$
$$Y(\cdot) := [\widehat{y}_1(\cdot)\ \widehat{y}_2(\cdot)\ \widehat{y}_3(\cdot)],$$
$$U(\cdot) := [\widehat{u}_1(\cdot)\ \widehat{u}_2(\cdot)\ \widehat{u}_3(\cdot)],$$

it follows from Prop. 5.3 that we can compute $N(\cdot)$ such that

$$\begin{bmatrix} X(\cdot) \\ U(\cdot) \end{bmatrix} N(\cdot) = I_3.$$

Then we solve

$$\begin{bmatrix} \dot{X}(\cdot) \\ Y(\cdot) \end{bmatrix} N(\cdot) = \begin{bmatrix} A(\cdot) & B(\cdot) \\ C(\cdot) & D(\cdot) \end{bmatrix}.$$

for $(A(\cdot), B(\cdot), C(\cdot), D(\cdot))$.

7.3.2 Results

We verify the state space representation obtained in the above section. We implement numerical integration algorithms to solve it using MATLAB. In the simulation, we applied the input $\widehat{u}_1(\cdot)$ to the identified model, and the corresponding output signal was generated. We compared the output signal generated in the simulation by the identified model to the one generated by the physical experiments, see Figure 7.8.

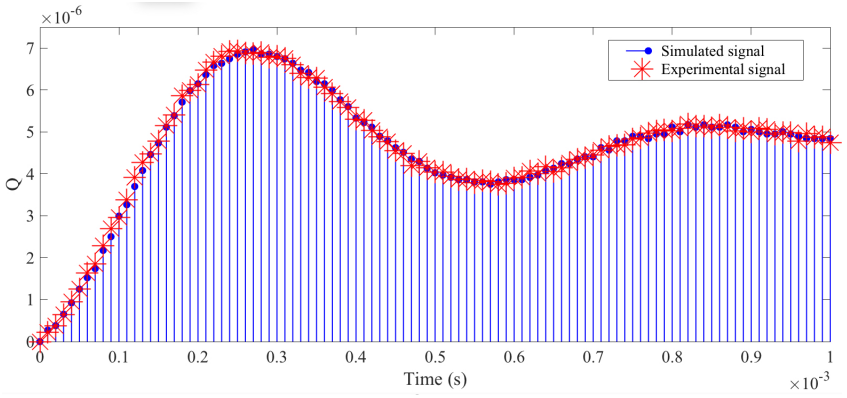


Figure 7.8: The charge of the capacitor from experimental data (red), and the charge of the capacitor from the identified model (blue).

We conclude that the experimental and the simulated data output, corresponding to the same input, show the same behaviour.

7.4 Summary

In this chapter the duality-based approach using RKHS was implemented on both a simulated and physical LTV self-adjoint system. We showed the procedure for the construction of functions using RKHS. The unfalsified state space model was solved using numerical integration algorithms. Finally, we compared the results obtained from the identified system to the ones obtained from the physical system and confirmed the accuracy of the results.

Chapter 8

Conclusions and future research

In this thesis, we developed a duality-based identification approach for LTV systems. LTV systems and RKHS are covered in Chapter 3. We used the notion of duality, covered in Chapter 4, to show the connection between the primal and the dual system. The duality-based approach is based on the fact that external structures, e.g. inputs and outputs, are reflected in the internal ones, e.g. states. We developed conditions under which one can identify a linear time-varying system using duality. We also introduced self-adjointness that provides advantages for identification. In the following we outline the contribution of this thesis.

In Chapter 4 we introduced a framework for the study of duality in linear time-varying systems. In the following, we summarize the main contributions of this framework:

- In Chapter 4, we introduced an intrinsic trajectory-based definition of the adjoint system; we also provided the connection between the system matrices of primal and dual system.
- In Proposition 4.6, we showed the existence of a non-unique factorization of the state trajectories. This supports our approach in the factorization process in Chapter 5.
- We introduced the concept of self-adjointness to study the conditions in which the system presents this property, see Prop. 6.1.

In Chapter 5, we developed a duality-based approach for the identification of linear time-varying system. The main contributions are the following:

- In Proposition 5.3, we use controllability of the system together with *genericity* to conclude that the system is identifiable.

- In Proposition 5.6, we introduced a factorization for the state trajectories. This factorization differs from the classical approaches, e.g. SVD-based or QR decomposition-based, in that we do not need to compute subspaces to extract the system matrices, which results in increasing the computational complexity of the algorithm; we instead partition the matrix in blocks such that the decomposition is based on submatrices.
- We developed an algorithm for the computation of the system matrices based on the factorization given in Prop.5.6.
- In Section 5.4, we introduced an algorithm for the identification of a linear time-varying self-adjoint system. This allows us to develop algorithms to identify Hamiltonian systems without the need of dual trajectories, i.e. the dual trajectories are considered to be the same as the primal ones by definition.

In Chapter 6, we develop a duality-based approach for the identification of Hamiltonian systems. Finally, in Chapter 7 we implemented the theory to a simulated and physical system. The validation of the results are provided showing the accurate predictions with data not used for the identification process.

The duality-based identification approach developed in this work provides some advantages over current identification methods for LTV systems. Our approach in particular does not need the assumption of prior knowledge of time variation properties, e.g. fast time-varying or slowly time-varying behaviours. This is a key element in identification, since ideally prior information of the to-be-identified system is not given at hand. Another useful property is that we use multiple responses for identification. This has been shown to increase the accuracy, since the identification approach is able to capture more dynamics than with single responses. This also allows the study of systems under different scenarios, which is an advantage in simulation and control.

Future research

The results presented in this thesis are preliminary for the application in relevant research modelling problems such as switched systems, or renewable energy systems. In the following we outline some future research directions based on the obtained results.

- *Lossy systems.* The identification of LTV systems using the duality-based approach require the availability of the dual trajectories. However, if the losses of a system are available for measurements, then the identification procedure can be implemented. This is because when one uses the supply rate $E(\cdot) = \frac{1}{2}\left(u(\tau)^\top y(\tau) + y(\tau)^\top u(\tau)\right)$, then the matrix $E(\cdot)$ becomes the sum of the dissipated and the storage energy. Then having the total energy and the dissipation,

we only need to extract the information of the storage energy that is related to a quadratic function of the state. The next objective in my research is to develop the duality-based approach considering that the losses of the system are available for measurement. This can be achieved using an energy-based description, such as Port-controlled Hamiltonian (PCH) systems, see Van der Schaft (2006). This approach underlines the modelling of systems in terms of the storage energy, dissipation, and interconnection structure. These will allow us to define conditions for identification based on energy-based modelling. Application of system identification approaches for power electronic systems where losses can be measured will be studied.

- *Lossless switched systems.* The modelling of lossless switched system as multi level boost converter using switch inductors has been presented in Mousa et al. (2010). However, identification for such system has not been addressed. The development of identification procedures for lossless switch system is the next stage of my research exploiting lossless properties as a condition for self-adjointness studied in this work. This will be translated as preliminary of duality-based system identification for switch system.

- *Dual trajectories.* In this framework, dual trajectories are required for the identification of linear time-varying systems in the nonself-adjoint case. However, dual trajectories are not available in real-life applications. In the LTI case, this has been addressed in Rapisarda and Antoulas (2015b) by constructing them from the primal trajectories. A future research direction includes the development of a method of finding conditions under one can construct the trajectories of the dual from the primal ones.

Appendix A

Background material

In this thesis we work with matrices of analytic functions and we define the rank of a matrix of analytic functions as follows.

Definition A.1. Let $M(\cdot) \in \mathcal{A}^{n\times m}$

1. The row (column) rank of $M(\cdot)$ is the maximal number of independent rows (columns).
2. $M(\cdot)$ is said to be full row (column) rank if its rank equals the number of rows (columns).

Note that $M(\cdot)$ is different from $M(t)$ where $M(t)$ is the matrix function $M(\cdot)$ evaluated at t. Moreover, note that if $M(\cdot) \in \mathcal{A}^{n\times m}$ then $M(t) \in \mathbb{R}^{n\times m}$.

Definition A.2. Let $M(\cdot) \in \mathcal{A}^{n\times m}$, and $N(\cdot) \in \mathcal{A}^{m\times n}$. $N(\cdot)$ is called the right inverse of $M(\cdot)$ if $M(\cdot)N(\cdot) = I_n$.

Definition A.3. Let $M(\cdot) \in \mathcal{A}^{n\times m}$, and $N(\cdot) \in \mathcal{A}^{n\times m}$. $N(\cdot)$ is called the left inverse of $M(\cdot)$ if $N(\cdot)M(\cdot) = I_m$.

Definition A.4. Let $f(\cdot), \cdots, f_n(\cdot)$ be a collection of polynomials, then an *algebraic variety* V in $\mathbb{R}^n$ is defined as

$$V := \{x \in \mathbb{R}^n \mid f_1(x) = \cdots = f_n(x) = 0\}.$$

If $V \neq \mathbb{R}^n$ then V is called proper algebraic variety.

Definition A.5. The *Wronskian* of a function vector $F(\cdot) = \text{col}(f_1(\cdot), f_2(\cdot), .., f_n(\cdot))$ whose functions are n-time differentiable is defined by

$$W(\cdot) := \begin{bmatrix} F(\cdot) & F(\cdot)^{(1)} & \cdots & F(\cdot)^{(n-1)} \end{bmatrix}.$$

Definition A.6. A function $f : \mathbb{R} \to \mathbb{R}^n$ is said to have *compact support* if the support of f is a compact set.

Definition A.7. A function $f(t)$ is *analytic* on a time interval if for every point t_a in the interval, it can be represented by a convergent power series on some subinterval containing t_a, i.e. $f(t)$ is analytic on an interval if it has a convergent Taylor series representation at each point in the interval.

Definition A.8. A *function space* is a space whose elements are functions, e.g. we can define a function space F as the set of all functions that maps $\mathbb{R}$ to $\mathbb{R}$, i.e. the set of all f such that $f : \mathbb{R} \to \mathbb{R}$.

Definition A.9. Let V be a vector space over $\mathbb{R}$. A real-valued function $\|\cdot\| : V \to [0, \infty)$ is called a *norm* on V if for all $f, g \in V$ and $\alpha \in \mathbb{R}$ the following conditions hold:

1. $\|f\| = 0$ if and only if $f = 0$
2. $\|\alpha f\| = |\alpha| \|f\|$
3. $\|f + g\| \leq \|f\| + \|g\|$.

$(V, \|\cdot\|)$, or simply V, is then called a normed vector space. Note that a norm on V induces a metric d on V defined by $d(f, g) = \|f - g\|$ for $f, g \in V$.

Definition A.10. A sequence $\{f_n\}$ of elements in a *normed vector space* V is convergent if there is an $f \in V$ such that

$$\lim_{n\to\infty} \|f_n - f\| = 0 \tag{A.1}$$

Definition A.11. A sequence $\{f_n\}$ of elements in a normed vector space V is Cauchy if for every $c > 0$ there is a positive integer N, depending on c, such that for every $m, n > N$ the following inequality holds

$$\|f_m - f_n\| < c. \tag{A.2}$$

Definition A.12. A normed vector space V is complete if every Cauchy sequence in V converges.

Definition A.13. Let V be a vector space over $\mathbb{R}$. A function $\langle \cdot, \cdot \rangle : V \times V \to \mathbb{R}$ is called an inner product on V if for all $f, g, h \in V$ and $\alpha \in \mathbb{R}$ the following conditions hold:

1. $\langle f + g, h \rangle = \langle f, h \rangle + \langle g, h \rangle$
2. $\langle \alpha f, h \rangle = \alpha \langle f, h \rangle$

3. $\langle f, g\rangle = \langle g, f\rangle$
4. $\langle f, f\rangle \geq 0$, and $\langle f, f\rangle = 0$ iff $f = 0$.

A vector space equipped with an inner product is called inner product space. We can always define a norm on V induced by the inner product as follows

$$\|f\| = \sqrt{\langle f, f\rangle} \tag{A.3}$$

with the metric on V given by

$$\|f - g\| = \sqrt{\langle f - g, f - g\rangle}. \tag{A.4}$$

Definition A.14. A complete inner product space $\mathcal{H}$ is a Hilbert space.

Operators

Definition A.15. Let X and Y be normed vector spaces over $\mathbb{R}$. A linear operator $L : X \to Y$ is an operator such that for all $x, y \in X$ and $\alpha \in \mathbb{R}$ it satisfies the following conditions:

1. $L(x + y) = L(x) + L(y)$
2. $L(\alpha x) = \alpha L(x)$.

Definition A.16. Let X and Y be normed spaces over $\mathbb{R}$, and let $L : X \to Y$ be a linear operator. The linear operator L is said to be bounded if there is an $\alpha \in \mathbb{R}$ such that $\forall x \in X$

$$\|L(x)\|_Y \leq \alpha \|x\|_X. \tag{A.5}$$

Note that the norm $\|L(x)\|_Y$ is on Y and $\|x\|_X$ is on X.

Theorem A.17. *Let X and Y be normed vector spaces. Let $L : X \to Y$ be a linear operator. Then*

- *L is continuous iff L is bounded*
- *If L is continuous at a single point, then L is continuous.*

Proof. See Kreyszig (1989), p. 97 proof of theorem 2.7-9. □

Functionals

Definition A.18. Let X be a vector space over K. A functional $\mathcal{F}$ on X is a function with domain $\mathcal{D}(\mathcal{F}) \subseteq X$ and an image in the field K

$$\mathcal{F} : \mathcal{D}(\mathcal{F}) \to K. \tag{A.6}$$

The terms *functional* and *form* will be used equivalently.

Definition A.19. Let X be a normed vector space. A linear functional $\mathcal{F} : X \to K$ is a functional such that for all $f, g \in X$ and $\alpha \in \mathbb{R}$ it satisfies

1. $\mathcal{F}(\alpha f) = \alpha \mathcal{F}(f)$,
2. $\mathcal{F}(f + g) = \mathcal{F}(f) + \mathcal{F}(g)$.

Definition A.20. Let X be a normed vector space. A linear functional $\mathcal{F}$ on X is said to be bounded if there exist a real number $c > 0$ such that for all $x \in X$

$$\|\mathcal{F}(x)\| \leq c\|x\|.$$

Definition A.21. Let X be a normed space over $\mathbb{R}$. The space X^* of all bounded linear functionals $\mathcal{F} : X \to \mathbb{R}$ is called the dual space of X.

Theorem A.22. *A linear functional $\mathcal{F}$ in a normed vector space is continuous iff it is bounded.*

References

Aronszajn, N. (1950). Theory of reproducing kernels. *Transactions of the American Mathematical Society*, 68(3):337–404.

Ba, T., Guan, X., Zhang, J. W., and Wang, S. (2016). Application of recursive subspace method in vehicle lateral dynamics model identification. *Mathematical Problems in Engineering*, 2016(1):15.

Chen, C.-T. (1967). Linear independence of analytic functions. *SIAM Journal on Applied Mathematics*, 15(5):1272–1274.

Crouch, P. and Van der Schaft, A. (1987). *Variational an Hamiltonian Control Systems*. Lecture Notes in Control and Information Sciences. Springer-Verlag New York.

Crouch, P. E., Lamnabhi-Lagarrigue, F., and der Schaft, A. J. V. (1995). Adjoint and hamiltonian input-output differential equations. *IEEE Transactions on Automatic Control*, 40(4):603–615.

de Almeida Rego, J. B., de Medeiros Martins, A., and de B. Costa, E. (2014). Deterministic system identification using rbf networks. *Mathematical Problems in Engineering*, 2014(1):1–10.

Dewilde, P. and der Veen, A.-J. V. (1998). *Time-Varying Systems and Computations*. Springer, US, 1 edition.

Heij, C. (1989). *Deterministic Identification of Dynamical Systems*. Springer-Verlag Berlin Heidelberg.

Huang, C. S., Hung, S. L., Su, W. C., and Wu, C. L. (2009). Identification of time-variant modal parameters using time-varying autoregressive with exogenous input and low-order polynomial function. *Computer-Aided Civil and Infrastructure Engineering*, 24(7):470–491.

Isermann, R. and Münchhof, M. (2014). *Identification of Dynamic Systems: An Introduction with Applications*. Springer Berlin Heidelberg.

Jhinaoui, A., Mevel, L., and Morlier, J. (2012). Subspace identification for linear periodically time-varying systems. *IFAC Proceedings Volumes*, 45(16):1282 – 1287.

Kailath, T. (1980). *Linear Systems*. Information and System Sciences Series. Prentice-Hall.

Kalman, R. (1960). Contributions to the theory of optimal control. *Bol. Soc. Mat. Mexicana*, 5.2:102–119.

Keesman, K. (2011). *System Identification: An Introduction*. Advanced Textbooks in Control and Signal Processing. Springer London.

Kreyszig, E. (1989). *Introductory Functional Analysis with Applications*. Wiley classics library. John Wiley & Sons.

Li, Y., l. Wei, H., and Billings, S. A. (2011). Identification of time-varying systems using multi-wavelet basis functions. *IEEE Transactions on Control Systems Technology*, 19(3):656–663.

Li, Y., Liu, Q., Tan, S.-R., and Chan, R. H. (2016). High-resolution time-frequency analysis of EEG signals using multiscale radial basis functions. *Neurocomputing*, 195:96 – 103.

Liu, K. (1997). Identification of linear time-varying systems. *Journal of Sound and Vibration*, 206(4):487 – 505.

Mousa, M., Ahmed, M., and Orabi, M. (2010). A switched inductor multilevel boost converter. In *2010 IEEE International Conference on Power and Energy,* 819−823.

Niedzwiecki, M. and Klaput, T. (2002). Fast recursive basis function estimators for identification of time-varying processes. *IEEE Transactions on Signal Processing*, 50(8):1925–1934.

Ohsumi, A. and Kawano, T. (2002). Subspace identification for a class of time-varying continuous-time stochastic systems via distribution-based approach. *15th IFAC World Congress*, 35(1):241 – 246.

Pillonetto, G., Dinuzzo, F., Chen, T., Nicolao, G. D., and Ljung, L. (2014). Kernel methods in system identification, machine learning and function estimation: A survey. *Automatica*, 50(3):657 – 682.

Rapisarda, P. and Antoulas, A. C. (2015a). Bilinear differential forms and the loewner framework for rational interpolation. In Belur, M. N., Camlibel, M. K., Rapisarda, P., and Scherpen, J. M., editors, *Mathematical Control Theory II: Behavioral Systems and Robust Control*. Springer-Verlag, 23−24.

Rapisarda, P. and Antoulas, A. C. (2015b). A duality perspective on loewner rational interpolation and state-space modelling of vector-exponential trajectories. In *2015 54th IEEE Conference on Decision and Control (CDC),* 2096−2100.

Rapisarda, P. and der Schaft, A. V. (2013). Identification and data-driven reduced-order modeling for linear conservative port- and self-adjoint hamiltonian systems. In *52nd IEEE Conference on Decision and Control,* 145−150.

Rapisarda, P. and Rao, S. (2013). Realization of lossless systems via constant matrix factorizations. *IEEE Transactions on Automatic Control*, 58(10):2632–2636.

Rapisarda, P. and Trentelman, H. (2011). Identification and data-driven model reduction of state-space representations of lossless and dissipative systems from noise-free data. *Automatica*, 47(8):1721 – 1728.

Rudolph, J. (1996). Duality in time-varying linear systems: A module theoretic approach. *Linear Algebra and its Applications*, 245:83 – 106.

Santos, C., Silva, B., and Yuan, J. (1998). Block sor methods for rank-deficient least-squares problems. *Journal of Computational and Applied Mathematics*, 100(1):1 – 9.

Schölkopf, B., Herbrich, R., and Smola, A. J. (2001). A generalized representer theorem. In Helmbold, D. and Williamson, B., editors, *Computational Learning Theory: 14th Annual Conference on Computational Learning Theory, COLT 2001 and 5th European Conference on Computational Learning Theory, EuroCOLT 2001 Amsterdam, The Netherlands, July 16–19, 2001 Proceedings*. Springer Berlin Heidelberg, 416−426.

Shawe-Taylor, J. and Cristianini, N. (2004). *Kernel Methods for Pattern Analysis*. Cambridge University Press, New York, USA.

Shi, Z. Y., Law, S. S., and Li, H. N. (2007). Subspace-based identification of linear time-varying system. *AIAA Journal*, 45(8):2042–2050.

Silverman, L. (1966). *Representation and Realization of Time-variable Linear Systems*. Columbia University New York, Dept. of Electrical Engineering, Technical report.

Silverman, L. M. and Meadows, H. E. (1967). Controllability and observability in time-variable linear systems. *SIAM Journal on Control*, 5(1):64–73.

Steinwart, I. and Christmann, A. (2008). *Support Vector Machines*. Information Science and Statistics. Springer New York.

Van der Schaft, A. (2006). Port-hamiltonian systems: an introductory survey. In Sanz-Sole, M., Soria, J., Varona, J., and Verdera, J., editors, *Proceedings of the International Congress of Mathematicians Vol. III*. European Mathematical Society Publishing House (EMS Ph), 3(2):1339−1365.

Van der Schaft, A. J. (1991). Duality for linear systems: External and state space characterization of the adjoint system. *Analysis of Controlled Dynamical Systems*, (1):393–403.

Van Der Veen, A., Deprettere, E. F., and Swindlehurst, A. L. (1993). Subspace-based signal analysis using singular value decomposition. *Proceedings of the IEEE*, 81(9):1277–1308.

Verhaegen, M. and Verdult, V. (2007). *Filtering and System Identification: A Least Squares Approach*. Cambridge University Press, New York, NY, USA, 1 edition.

Verhaegen, M. and Yu, X. (1995). A class of subspace model identification algorithms to identify periodically and arbitrarily time-varying systems. *Automatica*, 31(2):201 – 216.